Implementation of Dynamic Semiparametric Factor Model for Implied Volatilities

ABSCHLUSSARBEIT

zur Erlangung des akademischen Grades

Master of Science

(M.Sc.)

im Masterstudiengang Statistik

an der Wirtschaftwissenschaftlichen Fakultät

Humboldt-Universität zu Berlin

von

Szymon Borak

geboren am 03.05.1979 in Wrocław

Gutachter:

Prof. Dr. Wolfgang Härdle
Prof. Dr. Bernd Rönz

eingereicht am May 20, 2005

Implementation of Dynamic Semiparametric Factor Model for Implied Volatilities

A Master Thesis Presented

by

Szymon Borak

(189049)

to

Prof. Dr. Wolfgang Härdle

CASE - Center of Applied Statistics and Economics

Humboldt University, Berlin

in partial fulfillment of the requirements

for the degree of

Master of Science

Berlin, May 20, 2005

Declaration of Authorship

I hereby confirm that I have authored this master thesis independently and without use of others than the indicated resources. All passages, which are literally or in general matter taken out of publications or other resources, are marked as such.

Szymon Borak

Berlin, May 20, 2005

Abstract

Dynamic Semiparametric Factor Model (DSFM) is a convenient tool for analysis of implied volatility surfice (IVS). It offers dimension reduction of the IVS and can be therefore applied in hedging, prediction or risk mangement. However the estimation of the DSFM parameters is a complex procedure since it requires huge number of observation. Therefore the efficient implementation is a key issue for application possibilites of this model.

In this master thesis we discuss implementation issues of DSFM. We describe key features of the model and present its implementation in statistical computing enviroment XploRe.

Keywords: Dynamic Semiparametric Factor Model, Implied Volatility, Option Pricing

Contents

List of Figures

Notation and Abbreviations

K	strike price
C^{KO}	price of the call knock-out option
C	price of the plain vanilla call option
κ	moneyness
τ	time to maturity
$X_{i,j}$	exploratory variable containing moneyness and time to maturity
$Y_{i,j}$	observed log-implied volatility
$\widehat{m}$	basis function
$\widehat{\beta}_i$	loading time series
$\mathbf{Q}$	martingale measure
$\mathbf{P}$	market measure
$\widehat{\sigma}$	implied volatility
$\Phi(\cdot)$	is a cumulative distribution function of standard normal variable
S_t	asset price process
W_t	Wiener process
Z_t	compound Poisson process
BS	Black-Scholes
DSFM	Dynamic Semiparametric Factor Model
IV	Implied Volatility
IVS	Implied Volatility Surface
LV	Local Volatility
LVS	Local Volatility Surface
PDE	Partial Differential Equation
PCA	Principal Component Analysis
HP	Hedge Portfolio

1 Introduction

In economic modelling a problem with tradeoff between too complex models and too simple ones arises very often. On the one hand too complex modelling which analyze many aspects may lead to infeasible models, which despite their good fitting ability cannot be applied. On the other hand too simple models can miss some important features deviate much from the reality.

Recently the modelling can be more and more complex due to the development in computation technology. One may handle great quantity of high dimensional data and analyze several aspects simultaneously. However the optimal implementation is still the key issue.

In modern quantitative finance one has to handle with great quantity of more dimensional data too. The simple model cannot always explain the stylized facts which arise form the analysis of these data. Therefore more complex modelling approaches, which are still feasible, is being constantly proposed. The standard example is the implied volatility (IV). Both on the daily and intra-day level one may observe many option trades, which results in rich structure. Neglecting this structure like in Black-Scholes(BS) model may lead to pure forecasting performance. On the other hand too detailed analysis can induce too complex models for the applications.

In this thesis we discuss implementation of Dynamic Semiparemetric Factor Model (DSFM). The model can be successfully used for analysis of implied volatilities and we will focus on this particular application. It supports both enough complexity and can be easily tractable. However for the tractability of the model one needs still efficient estimation procedure. In the thesis we not only discuss the implementation issues but also extend the functionality of statistical package XploRe, which supports now the convenient way of handling the model.

The thesis is organized as follows. In Chapter 2 the overview of financial modelling for IV is presented. We present the popular BS model and define the IV concept. Some financial models which try to catch the IV structure are also presented. Chapter 3 focuses on presenting the DSFM. We formulate the model, present the estimation procedure and discuss some estimation's issues. In Chapter 4 the implementation issues are presented. This part recalls the numerical algorithms and discuss implementation

problems. The XploRe implementation is also described in details and efficiency study of the algorithm presented. Chapter 5 focusses on some application of the DSFM. The fit of the model to DAX options is considered, some simulation study and application to hedging is presented. We believe that due to optimal implementation this application can be efficiently studied in the future work.

2 Implied Volatilities

Recently implied volatility (IV) has become popular among practitioners to quote options prices. Due to its simplicity it gives an easy way to compare prices of options with different strike prices and different times to maturity. The idea is directly derived from the Black-Scholes (BS) formula, which is one of the most recognizable result in modern quantitative finance. The IV concept, however, appeared as a contradiction of the assumptions of BS model. There are constant attempts which try to removed its deficiencies by more complex modelling, which take into account the real IV behavior.

This chapter focusses on different aspects of the IV modelling. First we present BS formula for pricing European plain vanilla options. Then we introduce the concept of implied volatility and discuss the empirical facts of the option market which do not confirm the assumptions of BS model. In Section 2.3 and Section 2.4 some models consistent with non flat implied volatility surface are presented. The last Section focuses on models which try to catch the dynamic behavior of IVs.

2.1 Black-Scholes Formula

The work of Black and Scholes (1973) is one of the most recognized results in quantitative finance. The presented model assumes continuous trading on the time horizon $[0, T]$ and probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The filtration is defined by Wiener process W_t. The price of the tradable asset is a stochastic process given by stochastic differential equation:

$$dS_t = \mu S_t dt + \sigma S_t dW_t, \tag{2.1}$$

where the μ is the constant drift and σ is the volatility. The parameter μ describes the trend of the price evolution and σ the intensity of random deviations from the trend, which are caused by vibrations in price due to eg. temporary imbalance in supply and demand. On the stock market no transactions cost is assumed and buying or short selling all possible quantities of the asset is always possible. There exist

also the money market with equal rate for borrowing and lending r. The continuous compounded interest rate leads to the price of the zero coupon bond in time $t_0 = 0$ paying one unit in time t given by:

$$B_t = e^{-rt}.$$

The term structure of the interest rate is flat and r is constant on the time horizon $[0, T]$.

With Ito formula one can transform the price process to:

$$S_t = S_0 \exp\left\{(\mu - \frac{1}{2}\sigma^2)t + \sigma W_t\right\}. \tag{2.2}$$

This representation allows to induce the distribution of the price and many calculations become feasible.

In this framework Black and Scholes (1973) derived the price of the European plain vanilla options. The option is a financial contract which yields certain payments depending on the price of the underlying asset in specific time T. The simplest option is plain vanilla call option where the payment is given by:

$$\max(S_T - K, 0) = (S_T - K)^+.$$

The option pays $S_T - K$ units only if the price of the asset is greater than the certain level K, which is called strike price. The contract which pays:

$$\max(K - S_T, 0) = (K - S_T)^+$$

is called put plain vanilla option. The time, which remains to the final moment of final payment T (maturity), time to maturity: $\tau = T - t$, where t is current time point. Obviously the right to get payment defined by the option needs to cost initial premium, because the potential payment is always positive. The formula for calculating this price is the main result of Black and Scholes (1973).

To price the option non-arbitrage methodology is applied. Two strategies giving the same payment in time T should have the same value in time $t_0 = 0$. In formal mathematical language it is required that the discounted price process $e^{-rt}S_t$ has to be an martingale. The Girsanov's theorem says that there exists a measure $\mathbf{Q}$ equivalent to $\mathbf{P}$ under which the process

$$W_t^{\mathbf{Q}} = W_t + \frac{\mu - r}{\sigma} t$$

is **Q**-Wiener process. Only under the new arbitrage free probability measure **Q** the discounted price process $e^{-rt}S_t$ becomes a martingale:

$$de^{-rt}S_t = \sigma e^{-rt} S_t dW_t^{\mathbf{Q}}$$

and the dynamics of the asset is now given by:

$$dS_t = rS_t dt + \sigma S_t dW_t^{\mathbf{Q}}. \tag{2.3}$$

Using now (2.2), where μ is substituted with r, the density function of the S_T given the price $S_0 = s_0$ can be calculated since $S_T|S_0 = s_0$ has log-normal distribution.

Since the measure **Q** is unique the BS model gives the price of the call option:

$$C_t(S_t, K, r, \tau, \sigma) = S_t\Phi(d_+) - Ke^{-r\tau}\Phi(d_-) \tag{2.4}$$

where

$$d_\pm = \frac{\ln \frac{S_t}{K} + (r \pm \frac{1}{2}\sigma^2)}{\sigma\sqrt{\tau}}$$

and $\Phi(\cdot)$ is a cumulative distribution function of standard normal variable. The put option prices can be calculated from Put-Call parity:

$$P_t(S_t, K, r, \tau, \sigma) = C_t(S_t, K, r, \tau, \sigma) - S_t + Ke^{-r\tau}.$$

The five parameters of the option price in (2.4) can to grouped into three categories. Firstly S_t and r may directly obtained from the market data. Of course there is plethora of possible choices for r, since constant in time risk free interest rate, which reveals flat term structure, does not exist in practice. Secondly K and τ are specified in the option contract. While K is a fixed number τ changes deterministically with time by decreasing linearly to zero through the life time of the option. Finally σ is not observable or specified volatility parameter and has to be estimated from historical prices. It reflects the variability of the asset price. The bigger is uncertainty of the possible asset price change the higher is the call option price. The call option price an increasing function of σ so there exist one-to-one mapping between option price and the volatility.

2.2 Implied Volatility

In the growing financial markets the derivative markets were established. The plain vanilla options became regularly traded instruments. One may trade parallel many contracts with different specification of strike price and time to maturity. It means that many option prices can be observed at the same time.

Since (2.4) contains only one quantity that is not observed and the call option is the increasing function of volatility, then σ can be uniquely calculated by inverting BS formula. The value $\widehat{\sigma}$ that match observed option prices with (2.4) is called *implied volatility* (IV). Although there exists no direct formula for calculating IV from the market data, it can be computed efficiently with some numerical methods like bisection.

The surface (on day t) given by the mapping from strikes and from time to maturity τ:

$$(K, \tau) \to \widehat{\sigma}_t(K, \tau)$$

is called *implied volatility surface* (IVS). Note that although IVS is defined for all positive strikes and maturities it can be observed only on finite number of points.

A convenient way of presenting the IVS is to rewrite it as a function of a moneyness and time to maturity. The moneyness is κ is generally defined as:

$$\kappa = m(t, T, S_t, K, r).$$

where m is the increasing function in K. From now on we will consider a IVS as function of moneyness κ and time to maturity τ:

$$(\kappa, \tau) \to \widehat{\sigma}_t(\kappa, \tau).$$

We follow Fengler (2004) and set the moneyness to forward (or future) moneyness $\kappa = \frac{K}{e^{r\tau} S_t}$. The other possible choices are discussed in Hafner (2004).

The volatility parameter σ is assumed to be constant in the BS model. This assumption would be equivalent to flat IVS, which is not changing in time. However empirical findings show that IVS reveals a non-flat profile across moneyness (called "smile" or "smirk") and across time to maturity. Figure 5.5 presents typical IVs observed on January, 4th 1999. They clearly form smiles in moneyness direction and the curvature of the smile is different for each maturity.

In Figure 5.5 it is also visible the IVS is observed only in some limited number of points. Due to institutional conventions of the option market in one time only several maturities are traded. IVs form typical strings with common time to maturity but different strikes (moneyness). The string structure can be even better observed in

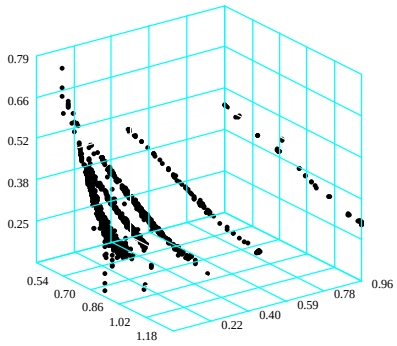


Figure 2.1: IVS ticks on January, 4th 1999

Figure 2.2, where 3-dimensional data is projected on time to maturity vs. moneyness plane. One may observe that near expiry there exist more strings than in the range with greater maturities.

Moreover strings move towards expiry as the time goes by. The time to maturity from today is not the time to maturity from tomorrow. Each day they shift slightly in direction of expiry (see Figure 2.3). Not only do they move but also change randomly the shape. All these effects make the modelling of IV and IVS a complex and challenging problem.

2.3 Alternative Financial Models

Section 2.1 presents the assumptions of BS model and derives the price of the European call option. Section 2.2 shows some empirical facts which contradict the BS assumptions. Despite the deficiencies BS model is popular due to its intuitive simplic-

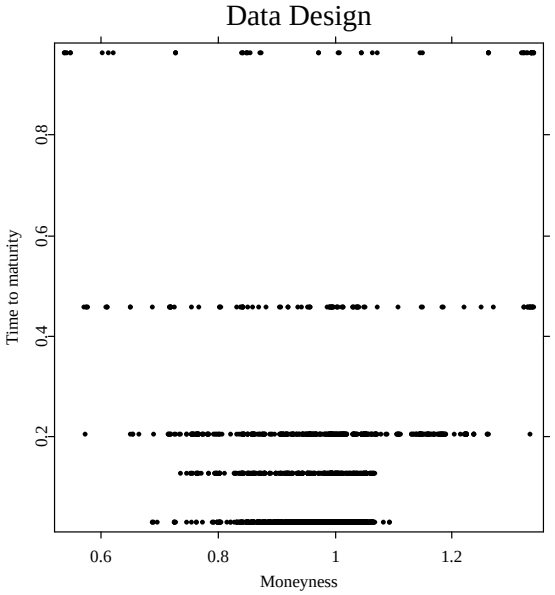


Figure 2.2: Data design on January, 4th 1999

ity. However for some financial applications BS simplification may result in significant inaccuracy. Standard example is pricing the exotic options when σ has to taken from the market and different σ leads to different prices. To overcome the problem with non flat IVS more complex financial models can be considered, which assume different stochastic behavior for the underlying. Among many models this section presents three particular models: Merton jump diffusion model, Heston stochastic volatility model and Bates stochastic volatility with jumps model.

2.3.1 Merton Model

If an important piece of information about the company becomes public it may cause a sudden change in the company's stock price. The information usually comes at a random time and the size of its impact on the stock price may be treated as a random variable. To cope with these observations Merton (1976) proposed a model that allows discontinuous trajectories of asset prices. The model extends (2.1) by adding jumps

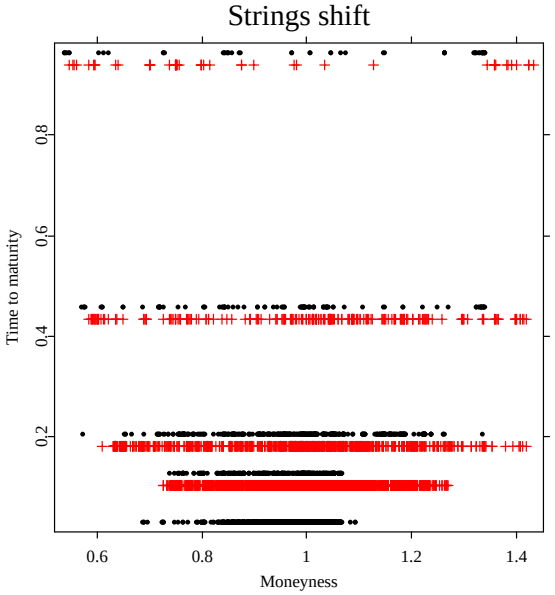


Figure 2.3: IV strings on January, 4th 1999 (points) and on January, 13th 1999 (crosses).

to the stock price dynamics:

$$\frac{dS_t}{S_t} = rdt + \sigma dW_t + dZ_t, \tag{2.5}$$

where Z_t is a compound Poisson process with a log-normal distribution of jump sizes. The jumps follow a (homogeneous) Poisson process N_t with intensity λ, which is independent of W_t. The log-jump sizes $Y_i \sim N(\mu, \delta^2)$ are i.i.d random variables with mean μ and variance δ^2, which are independent of both N_t and W_t.

The model becomes incomplete which means that there are many possible ways to choose a risk-neutral measure such that the discounted price process is a martingale. Merton proposed to change the drift of the Wiener process and to leave the other ingredients unchanged. The asset price dynamics is then given by:

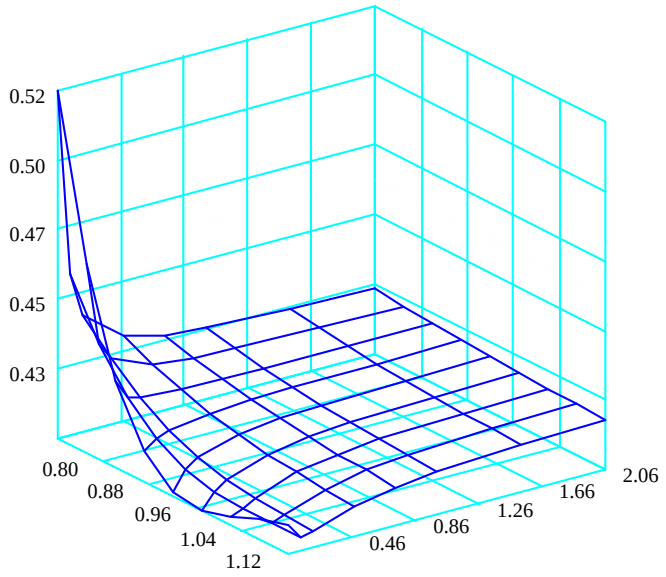


Figure 2.4: Implied volatility surface of the Merton model for $\mu^M = 0.046$, $\sigma = 0.15$, $\lambda = 0.5$, $\delta = 0.2$, and $m = -0.243$.

$$S_t = S_0 \exp\left(\mu^M t + \sigma W_t + \sum_{i=1}^{N_t} Y_i\right),$$

where $\mu^M = r - \sigma^2 - \lambda\{\exp(\mu + \frac{\delta^2}{2}) - 1\}$. Jump components add mass to the tails of the returns distribution. Increasing δ adds mass to both tails, while a negative/positive μ implies relatively more mass in the left/right tail.

The Merton model not only propose more realistic dynamics of the asset price but also generate non-constant IVS. The IVS obtained from this model is presented in Figure 2.4.

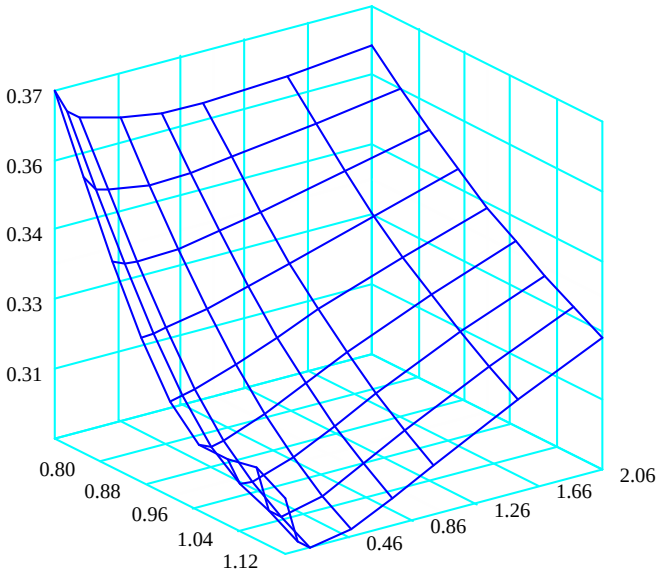


Figure 2.5: Implied volatility surface of the Heston model for $\xi = 1.0$, $\theta = 0.15$, $\sigma = 0.5$, $v_0 = 0.1$, and $\rho = -0.5$.

2.3.2 Heston Model

Another possible modification of (2.1) is to substitute the constant volatility parameter σ with a stochastic process. This leads to the so-called "stochastic volatility" models, where the price dynamics is driven by:

$$\frac{dS_t}{S_t} = rdt + \sqrt{v_t}dW_t,$$

where v_t is another unobservable stochastic process. There are many possible ways of choosing the variance process v_t. Hull and White (1987) proposed to use geometric Brownian motion:

$$\frac{dv_t}{v_t} = c_1 dt + c_2 dW_t. \tag{2.6}$$

However, geometric Brownian motion tends to increase exponentially which is an undesirable property for volatility. Volatility exhibits rather a mean reverting behavior.

Therefore a model based on an Ornstein-Uhlenbeck-type process:

$$dv_t = \xi(\theta - v_t)dt + \beta dW_t, \tag{2.7}$$

was suggested by Stein and Stein (1991). This process, however, admits negative values of the variance v_t.

These deficiencies were eliminated in a stochastic volatility model introduced by Heston (1993):

$$\begin{aligned} \frac{dS_t}{S_t} &= r dt + \sqrt{v_t} dW_t^{(1)}, \\ dv_t &= \xi(\theta - v_t)dt + \sigma\sqrt{v_t} dW_t^{(2)}, \end{aligned} \tag{2.8}$$

where the two Brownian components $W_t^{(1)}$ and $W_t^{(2)}$ are correlated with rate ρ:

$$\mathrm{Cov}\left(dW_t^{(1)}, dW_t^{(2)}\right) = \rho dt. \tag{2.9}$$

The term $\sqrt{v_t}$ in equation (2.9) simply ensures positive volatility. When the process touches the zero bound the stochastic part becomes zero and the non-stochastic part will push it up.

Parameter ξ measures the speed of mean reversion, θ is the average level of volatility and σ is the volatility of volatility. In (2.8) the correlation ρ is typically negative, what is known as the "leverage effect". Empirical studies of the financial returns confirm that volatility is negatively correlated with the returns.

The risk neutral dynamics is given in a similar way as in the BS model. For the logarithm of the asset price process $X_t = \ln \frac{S_t}{S_0}$ one obtains the equation:

$$dX_t = \left(r - \frac{1}{2}v_t\right) dt + \sqrt{v_t} dW_t^{(1)}.$$

Figure 2.5 presents non-constant IVS generated from Heston model with arbitrary chosen parameters: $\xi = 1.0$, $\theta = 0.15$, $\sigma = 0.5$, $v_0 = 0.1$, and $\rho = -0.5$.

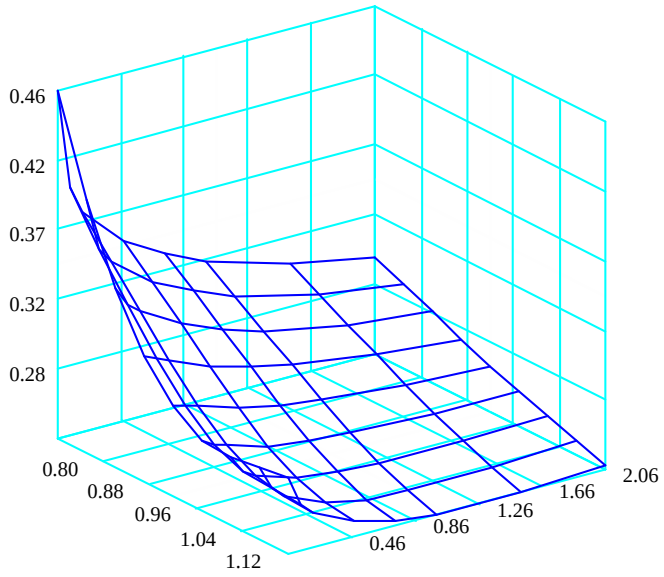


Figure 2.6: Implied volatility surface of the Bates model for $\lambda = 0.5,\ \delta = 0.2,\ \overline{k} = -0.1,\ \xi = 1.0,\ \theta = 0.15,\ \rho = -0.5,\ \sigma = 0.5$ and $v_0 = 0.1$.

2.3.3 Bates Model

The Merton and Heston approaches were combined by Bates (1996), who proposed a model with stochastic volatility and jumps:

$$\begin{aligned} \frac{dS_t}{S_t} &= r dt + \sqrt{v_t} dW_t^{(1)} + dZ_t, \\ dv_t &= \xi(\theta - v_t)dt + \sigma\sqrt{v_t} dW_t^{(2)}, \\ \mathrm{Cov}(dW_t^{(1)}, dW_t^{(2)}) &= \rho\, dt. \end{aligned} \tag{2.10}$$

As in (2.5) Z_t is a compound Poisson process with intensity λ and log-normal distribution of jump sizes independent of $W_t^{(1)}$ (and $W_t^{(2)}$). If J denotes the jump size then $\ln(1+J) \sim N(\ln(1+\overline{k}) - \frac{1}{2}\delta^2, \delta^2)$ for some $\overline{k}$. Under the risk neutral probability one

obtains the equation for the logarithm of the asset price:

$$dX_t = (r - \lambda\overline{k} - \frac{1}{2}v_t)dt + \sqrt{v_t}dW_t^{(1)} + \tilde{Z}_t,$$

where $\tilde{Z}_t$ is a compound Poisson process with normal distribution of jump magnitudes.

Similarly to the Merton and Heston models the Bates model also yields non-constant IVS. This model has eight parameters and out of the three presented models can mimic the IVS most precisely. However one needs to pay the price of more difficult calibration of the parameters. The calibration issues of the three models is presented in Detlefsen (2005). The fast and efficient method for calculating option prices for the three models is discussed in Borak et al. (2004).

2.4 Local Volatility Model

The models of the previous section modified the dynamics of the asset by imposing another degrees of freedom like jumps, stochastic process for volatility or both. Similarly to BS model the option prices depend on unknown parameters. However the number of parameters is greater than one, which allows to reproduce non-flat IVS.

Another possibility to model non-constant volatility surface is given by so called local volatility model (LV). The asset price follows the equation:

$$dS_t = rS_t dt + \sigma(S_t, t)S_t dW_t, \tag{2.11}$$

where $\sigma(S_t, t)$ is a deterministic function of the asset value S_t and time t. The function $\sigma(S_t, t)$ has two arguments so one may consider similarly to IVS local volatility surface (LVS). It is given by mapping $(S_t, t) \to \sigma(S_t, t)$.

The strength of LV model is in its ability to yields "smile consistent" option prices. The LVS can be obtained from the observed market prices by analytical transformation given in Dupire (1994) in following way:

$$\sigma(S_t, t) = \sqrt{2\frac{\frac{C_t(K,T)}{\partial T} + rK\frac{C_t(K,T)}{\partial K}}{K^2\frac{\partial^2 C_t(K,T)}{\partial K^2}}}. \tag{2.12}$$

Other advantage is fast pricing algorithm. To get prices in this model generalized BS partial differential equation (PDE):

$$\frac{C_t(K,T)}{\partial T} - \frac{1}{2}K^2\sigma^2(K,T)\frac{\partial^2 C_t(K,T)}{\partial K^2} + rK\frac{C_t(K,T)}{\partial K} = 0 \tag{2.13}$$

has to be solved numerically on a discrete grid. The existing algorithms, however, give the solution fast which simultaneously yield vanilla option prices for different strikes and different maturities.

2.5 Models of Implied Volatility Dynamics

A drawback of the more sophisticated models is the failure to correctly describe the dynamics of the IVS. This can be inferred from frequent recalibration of the model. Therefore models studying dynamic behavior of the IVS were proposed.

In modelling the dynamics of the IVS one face the problem of high complexity of the IV data. During one day simultaneously several maturities and levels of moneyness are observed. Therefore the vital stand of research was focused on the low-dimensional representation of the IVS. In this case principal component analysis (PCA), which is a standard method of extracting the most informative source of variation in multivariate systems, play a crucial role. PCA can be applied to moneyness, time structure or whole surfaces.

In this section we skim the dynamics models for the IVS. They offer low-dimensional representation and extract the factors of variation. We join the model performing PCA on the moneyness, PCA on the term structure and model extracting two dimensional functional factors.

2.5.1 PCA on the Moneyness

The string structure of the IV data results in problems with unification of the observations. The time to maturity from today of some specific option is not the time to maturity of the same option tomorrow. The analysis of IVs only with the same moneyness and time to maturity characteristics leads to even more complex structure and may result in lack of sufficient number of data, since the option with one month maturity will appear next time in one month time.

To overcome this problem Skiadopoulos et al. (1999) proposed to group the IVs into some buckets with similar moneyness and time to maturity. The moneyness space is divided into subintervals with $\kappa_1 < ... < \kappa_{N_\kappa}$. Similarly time to maturity is split in subintervals with $\tau_1 < ... < \tau_{N_\tau}$. Each observation belongs to one of the classes $[\kappa_i, \kappa_{i+1}) \times [\tau_i, \tau_{i+1})$. Then the PCA is performed on each maturity bucket $[\tau_i, \tau_{i+1})$. The smile is represented as a multivariate observation, where each coordinate is a observed IV from the range $[\kappa_i, \kappa_{i+1})$.

The dynamics of the smile is given by first p factors which explain the specified amount

of variance. The similar procedure may be redone for the whole IVS. For the bigger maturities however some observation can be missing and drastic enlargement of the buckets may be necessary. It may however influence the explanatory power of the model since some important features of the IVS dynamics could lost.

In order to obtain multivariate observation for the specific time to maturity and moneyness one may smooth the data on the grid. This approach was prosed by Fengler et al. (2003). Then for each maturity $\tau_1, ..., \tau_{N_\tau}$ the PCA can be performed on the multivariate observations with coordinates $\kappa_1, ..., \kappa_{N_\kappa}$. Instead of IV the log returns can be analyze as well. For dynamics of the whole IVS Fengler et al. (2003) propose the common PCA. The covariance matrix for each maturity Ψ_τ can be decompose:

$$\Psi_{\tau_i} = \Gamma \Lambda_{\tau_i} \Gamma^\top$$

where Γ is a matrix of eigenvectors and Λ_{τ_i} is diagonal matrix containing eigenvalues. The matrix Γ is assumed to be constant for each maturity τ_i and the only difference is given in Λ_{τ_i}.

2.5.2 PCA on the Term Structure

The PCA can be applied not only to the moneyness but also to the term structure. In extracting the single observation of the term structure for the fix moneyness κ once more one need to face the problem with moving strings. The standard approach is to consider fixed maturities eg. 1M, 2M, 3M etc. and calculate the IV as a linear interpolation between the observed IV strings.

Avellaneda and Zhu (1997) propose PCA model for the term structure of FX options. The term structure is reconstruct on some specified maturities $\tau_1, ..., \tau_{N_\tau}$. The observation in time t is then:

$$(\widehat{\sigma}(\tau_1), ..., \widehat{\sigma}(\tau_{N_\tau})).$$

The covariance matrix in PCA can be constructed not only from the data itself but also from log-returns. Avellaneda and Zhu (1997) analyze the matrix $S = s_{i,j}$ defined as

$$s_{i,j} = \frac{1}{T-1} \sum_t (\log \widehat{\sigma}_{t+1}(\tau_i) - \log \widehat{\sigma}_t(\tau_j))(\log \widehat{\sigma}_{t+1}(\tau_j) - \log \widehat{\sigma}_t(\tau_i)).$$

The log returns exclude the negative volatilities and show stationary behavior. After obtaining statistically uncorrelated principal components the log returns can be

modelled by:

$$\log \widehat{\sigma}_{t+1} - \log \widehat{\sigma}_t = \sum_{j=1}^{N_\tau} w_{j,t} PC_j$$

which lead to the IV term structure model:

$$\widehat{\sigma}_{t+1}(\tau_i) = \widehat{\sigma}_t \exp\left\{ \sum_{j=1}^{N_\tau} w_{j,t} PC_j^{(i)} \right\}$$

where $w_{j,t}$ are time dependent weights and $PC_j^{(i)}$ is the i-th coordinate of the principal component.

Fengler et al. (2002) perform the PCA for ATM options. Similarly to the approach of Section 2.5.1 the common PCA is employed to recover the structure of the whole IVS.

2.5.3 Dynamic Factor Models

The PCA analysis on the moneyness or the term structure respond only to part of the IVS dynamics. The dynamics of the whole IVS can be described by common PCA, since each slice of moneyness or term structure exhibit similar behavior. Another approach leads to the two-dimensional functional factors. The IVS is decomposed to small number of surfaces which operate on moneyness vs. time to maturity plane.

The functional PCA approach is presented in Cont and da Fonseca (2002). First the IVS estimates $\widehat{\sigma}_t^{NW}(\kappa, \tau)$ are obtained from kernel Nadaraya-Watson estimator and the log-return surfaces $\Delta \log \widehat{\sigma}_t^{NW}(\kappa, \tau) = \widehat{\sigma}_t^{NW}(\kappa, \tau) - \widehat{\sigma}_{t-1}^{NW}(\kappa, \tau)$ calculated. Then the Karhunen-Loéve decomposition, which a generalization of PCA to higher dimensional random fields, is applied to $\Delta \log \widehat{\sigma}_t^{NW}(\kappa, \tau)$. As a result one obtains the dynamic model for IVS

$$\widehat{\sigma}_t(\kappa, \tau) = \widehat{\sigma}_0(\kappa, \tau) \exp\left(\sum_l w_{t,l} g_l \right) \tag{2.14}$$

where the $w_{t,l}$ are time dependent loadings and g_l two-dimensional factor functions. The initial IVS is denoted by $\widehat{\sigma}_0$.

In the model presented above nonparametric estimation is applied and afterwards the dimension reduction to the estimates is given. However one may proceed in counter-order. Hafner (2004) propose first to reduce dimension to small number of factor functions and fit the model to this functions. The model is purely parametric since the function are not estimated but given a priori. The fit is done by least square minimization and the proposed functions are:

$$\begin{aligned} h_1(\kappa,\tau) &= 1 \\ h_2(\kappa,\tau) &= \log(\kappa) \\ h_3(\kappa,\tau) &= \{\log(\kappa)\}^2 \\ h_4(\kappa,\tau) &= \log(1+\tau) \\ h_5(\kappa,\tau) &= \log(\kappa)\log(1+\tau) \\ h_6(\kappa,\tau) &= \{\log(\kappa)\}^2\log(1+\tau). \end{aligned}$$

The model for the IVS dynamics has then the regression form:

$$\widehat{\sigma}_t(\kappa,\tau) = \sum_{l=1}^{6} \beta_{i,l} h_l.$$

The empirical analysis finds stable and strong relation between β_2 and β_5, and also β_3 and β_6. The factor loading β_5 is then substituted with $\varrho_1\beta_2$ and β_6 with $\varrho_2\beta_3$. For the options on DAX Hafner (2004) estimated with linear regression $\widehat{\varrho_1} = -1.6977$ and $\widehat{\varrho_2} = -3.3768$. Thus the model can be rewritten as four factor model:

$$\widehat{\sigma}_t(\kappa,\tau) = \beta_1+\beta_2\log(\kappa)\{1+\varrho_1\log(1+\tau)\}+\beta_3\{\log(\kappa)\}^2\{1+\varrho_1\log(1+\tau)\}+\beta_4\log(1+\tau) \tag{2.15}$$

3 Dynamic Semiparametric Factor Model

The models presented in the previous chapter which try to catch the dynamic behavior of the IVS disregard the specific string structure. The non-observable maturities are fitted on the particular day regardless of the observations from other days. This approach however may miss the important features of the IVS dynamics.

The DSFM proposed by Fengler et al. (2005) successfully cope with the generated data design. It offers flexible modelling for fitting, dimension reduction and explanation of dynamic behavior. This chapter focuses on presenting the DSFM. First we formulate the model and afterwards the estimation procedure is described in details. In Section 3.3 we show how the final solution is selected out of many equivalent solutions and in Section 3.4 we present the criteria for selecting the size of the model and bandwidths. In Sections 3.5 and 3.6 we discuss problems which arise in the estimation procedure due to the degenerated data design.

3.1 Model Formulation

The DSFM belongs to the class of models presented in Section 2.5.3. The IVS is assumed to be a weighted sum of the functional factors and the dynamics is explained by the stochastic behavior of the loadings. Contrary to the other models it simultaneously estimate the factor functions and fits the surface.

Let $Y_{i,j}$ be the log-implied volatility observed on a particular day. The index i is the number of the day, while the total number of days is denoted by I $(i = 1, ..., I)$. The index j represents an intra-day trade on day i and the number of trades on that day is J_i $(j = 1, ...J_i)$. Let $X_{i,j}$ be a two-dimensional variable containing moneyness $\kappa_{i,j}$ and maturity $\tau_{i,j}$. Among many moneyness settings we define it as $\kappa_{i,j} = \frac{K_{i,j}}{F_{t_i}}$, where $K_{i,j}$ is a strike and F_{t_i} the underlying futures price at time t_i. The DSFM regress $Y_{i,j}$

on $X_{i,j}$ by:

$$Y_{i,j} = m_0(X_{i,j}) + \sum_{l=1}^{L} \beta_{i,l} m_l(X_{i,j}), \tag{3.1}$$

where m_0 is an invariant basis function, m_l $(l = 1, ...L)$ are the 'dynamic' basis functions and $\beta_{i,l}$ are the factor weights depending on time i.

3.2 Estimation

The estimates $\widehat{\beta}_{i,l}$ and $\widehat{m}_l$ are obtained by minimizing the following least squares criterion $(\beta_{i,0} = 1)$:

$$\sum_{i=1}^{I} \sum_{j=1}^{J_i} \int \left\{ Y_{i,j} - \sum_{l=0}^{L} \widehat{\beta}_{i,l} \widehat{m}_l(u) \right\}^2 K_h(u - X_{i,j})\, du, \tag{3.2}$$

where K_h denotes two-dimension kernel function. The possible choice for two-dimension the kernel is a product of one dimension kernels $K_h(u) = k_{h_1}(u_1) \times k_{h_2}(u_2)$, where $h = (h_1, h_2)^\top$ are bandwidths and $k_h(v) = k(h^{-1}v)/h$ is a one dimensional kernel function.

The minimization procedure search through all functions $\widehat{m}_l : \mathbb{R}^2 \longrightarrow \mathbb{R}$ $(l = 0, ..., L)$ and time series $\widehat{\beta}_{i,l} \in \mathbb{R}$ $(i = 1, ..., I; l = 1, ..., L)$.

When $L = 0$ the procedure reduce to Nadaraya-Watson estimate based on the pooled sample of all days. This would neglect the dynamic structure yielding one estimate for all days. When additionally the sample length is limited to the one day $(I = 1)$ then simply Nadaraya-Watson estimate of the IVS of that particular day is obtained.

To calculate the estimates an iterative procedure is applied. First we introduce the following notation for $1 \leq i \leq I$:

$$\widehat{p}_i(u) = \frac{1}{J_i} \sum_{j=1}^{J_i} K_h(u - X_{i,j}), \tag{3.3}$$

$$\widehat{q}_i(u) = \frac{1}{J_i} \sum_{j=1}^{J_i} K_h(u - X_{i,j}) Y_{i,j}. \tag{3.4}$$

We denote by $\widehat{m}^{(r)} = (\widehat{m}_0^{(r)}, ..., \widehat{m}_L^{(r)})^\top$ the estimate of the basis functions and $\widehat{\beta}_i^{(r)} = (\widehat{\beta}_{i,l}^{(r)}, ..., \widehat{\beta}_{i,L}^{(r)})^\top$ the factor loadings on the day i after r iterations. By replacing each function $\widehat{m}_l$ in (3.2) by $\widehat{m}_l + \delta g$ with arbitrary function g and taking derivatives with respect to δ one obtains:

$$\frac{d}{d\delta} \sum_{i=1}^{I} \sum_{j=1}^{J_i} \int \left\{ Y_{i,j} - \sum_{l=0}^{L} \widehat{\beta}_{i,l} \widehat{m}_l(u) - \widehat{\beta}_{i,l'} \delta g \right\}^2 K_h(u - X_{i,j}) \, du =$$

$$2 \int \sum_{i=1}^{I} \sum_{j=1}^{J_i} \left\{ Y_{i,j} - \sum_{l=0}^{L} \widehat{\beta}_{i,l} \widehat{m}_l(u) - \widehat{\beta}_{i,l'} \delta g \right\} \widehat{\beta}_{i,l'} g K_h(u - X_{i,j}) \, du = 0. \tag{3.5}$$

Since the minimum is obtained for $\delta = 0$ and for any function g the integral in (3.5) is 0 if:

$$\sum_{i=1}^{I} \sum_{j=1}^{J_i} \left\{ Y_{i,j} - \sum_{l=0}^{L} \widehat{\beta}_{i,l} \widehat{m}_l(X_{i,j}) \right\} \widehat{\beta}_{i,l'} K_h(u - X_{i,j}) = 0. \tag{3.6}$$

Rearranging terms in (3.6) and plugging in (3.3)-(3.4) yields:

$$\sum_{i=1}^{I} J_i \widehat{\beta}_{i,l'} \widehat{q}_i(u) = \sum_{i=1}^{I} J_i \sum_{l=0}^{L} \widehat{\beta}_{i,l'} \widehat{\beta}_{i,l} \widehat{p}_i(u) \widehat{m}_l(u), \tag{3.7}$$

for $0 \leq l' \leq L$. In fact (3.7) is a set of $L+1$ equations. Define the matrix $B^{(r)}(u)$ and vector $Q^{(r)}(u)$ by their elements:

$$\left(B^{(r)}(u)\right)_{l,l'} = \sum_{i=1}^{I} J_i \widehat{\beta}_{i,l'}^{(r-1)} \widehat{\beta}_{i,l}^{(r-1)} \widehat{p}_i(u), \tag{3.8}$$

$$\left(Q^{(r)}(u)\right)_l = \sum_{i=1}^{I} J_i \widehat{\beta}_{i,l}^{(r-1)} \widehat{q}_i(u). \tag{3.9}$$

Thus (3.7) is equivalent to:

$$B^{(r)}(u) \widehat{m}^{(r)}(u) = Q^{(r)}(u) \tag{3.10}$$

which yields the estimate of $\widehat{m}^{(r)}(u)$ in the r-th iteration.

A similar idea has to be applied to update $\widehat{\beta}_i^{(r)}$. Replacing $\widehat{\beta}_{i,l}$ by $\widehat{\beta}_{i,l} + \delta$ in (3.2) and taking once more the derivative with respect to δ yields:

$$\sum_{j=1}^{J_i} \int \left\{ Y_{i,j} - \sum_{l=0}^{L} \widehat{\beta}_{i,l} \widehat{m}_l(X_{i,j}) \right\} \widehat{m}_{l'}(u) K_h(u - X_{i,j}) du = 0, \tag{3.11}$$

which leads to:

$$\int \widehat{q}_i(u) \widehat{m}_{l'}(u)\ du = \sum_{l=0}^{L} \widehat{\beta}_{i,l} \int \widehat{p}_i(u) \widehat{m}_{l'}(u) \widehat{m}_l(u)\ du, \tag{3.12}$$

for $1 \leq l' \leq L$. The formula (3.12) is now a system of L equations. Define the matrix $M^{(r)}(i)$ and the vector $S^{(r)}(i)$ by their elements:

$$\left(M^{(r)}(i)\right)_{l,l'} = \int \widehat{p}_i(u) \widehat{m}_{l'}(u) \widehat{m}_l(u)\ du, \tag{3.13}$$

$$\left(S^{(r)}(i)\right)_l = \int \widehat{q}_i(u) \widehat{m}_l(u)\ du - \int \widehat{p}_i(u) \widehat{m}_0(u) \widehat{m}_l(u)\ du. \tag{3.14}$$

An estimate of $\widehat{\beta}_i^{(r)}$ is thus given by solving:

$$M^{(r)}(i) \widehat{\beta}_i^{(r)} = S^{(r)}(i). \tag{3.15}$$

The algorithm stops when only minor changes occur:

$$\sum_{i=1}^{I} \int \left(\sum_{l=0}^{L} \widehat{\beta}_{i,l}^{(r)} \widehat{m}_l^{(r)}(u) - \widehat{\beta}_{i,l}^{(r-1)} \widehat{m}_l^{(r-1)}(u) \right)^2 du \leq \epsilon \tag{3.16}$$

for some small ϵ. Obviously one needs to set initial values of $\widehat{\beta}_i^{(0)}$ in order to start the algorithm.

3.3 Orthogonalization

The estimates $\widehat{m} = (\widehat{m}_1, ..., \widehat{m}_L)^\top$ of the basis functions are not uniquely defined. They can be replaced by functions that span the same affine space. Define $\widehat{p}(u) =$

$\frac{1}{I}\sum_{i=1}^{I}\widehat{p}_i(u)$ and the $L \times L$ matrix Γ by its elements

$$\Gamma_{l,l'} = \int \widehat{m}_l(u)\widehat{m}_{l'}(u)\widehat{p}(u)du.$$

The estimates $\widehat{m}$ are replaced by new functions $\widehat{m}^{new} = (\widehat{m}_1^{new}, ..., \widehat{m}_L^{new})^\top$:

$$\widehat{m}_0^{new} = \widehat{m}_0 - \gamma^\top \Gamma^{-1}\widehat{m} \tag{3.17}$$
$$\widehat{m}^{new} = \Gamma^{-1/2}\widehat{m} \tag{3.18}$$

such that they are now orthogonal in the $L^2(\widehat{p})$ space. The loading time series estimates $\widehat{\beta}_i = (\widehat{\beta}_{i,1}, ..., \widehat{\beta}_{i,L})^\top$ need to be substituted by:

$$\widehat{\beta}_i^{new} = \Gamma^{1/2}(\widehat{\beta}_i + \Gamma^{-1}\gamma), \tag{3.19}$$

where γ is $(L \times 1)$ vector with $\gamma_l = \int \widehat{m}_0(u)\widehat{m}_l(u)\widehat{p}(u)du$.

The next step is to choose an orthogonal basis such that for each $w = 1, ..., L$ the achieved explanation of the partial sum:

$$m_0(u) + \sum_{l=1}^{w} \beta_{i,l} m_l(u)$$

is maximal. One proceed as in PCA. First define matrix B with $B_{l,l'} = \sum_{i=1}^{I} \widehat{\beta}_{i,l}\widehat{\beta}_{i,l'}$ and $Z = (z_1, ..., z_L)$ where z_1,...,z_L are the eigenvectors of B. Then replace $\widehat{m}$ by $\widehat{m}^{new} = Z^\top \widehat{m}$ and $\widehat{\beta}_i$ by $\widehat{\beta}_i^{new} = Z^\top \widehat{\beta}_i$.

The orthonormal basis $\widehat{m}_1, ..., \widehat{m}_L$ is chosen such that $\sum_{i=1}^{I} \widehat{\beta}_{i,1}^2$ is maximal and given $\widehat{\beta}_{i,1}, \widehat{m}_0, \widehat{m}_1$ the quantity $\sum_{i=1}^{I} \widehat{\beta}_{i,2}^2$ is maximal and so forth.

3.4 Model selection

For the choice of the model size the residual sum of squares is calculated:

$$RV(L) = \frac{\sum_i \sum_j \left\{ Y_{i,j} - \sum_{l=0}^{L} \widehat{\beta}_{i,l}\widehat{m}_l(X_{i,j}) \right\}^2}{\sum_i \sum_j (Y_{i,j} - \overline{Y})^2}, \tag{3.20}$$

where $\overline{Y}$ is the overall mean of the observation. One may increase the parameter L until the explained variance $1 - RV(L)$ is sufficiently high. However if the model was

fitted for L dynamic functions, the new fit for the size $L+1$ requires repeating of almost entire procedure.

For the data-driven choice of bandwidths we take like Fengler et al. (2005) a weighted AIC. For the weight function w one needs to minimize:

$$\frac{1}{N}\sum_{i,j}\left\{Y_{i,j}-\sum_{l=0}^{L}\widehat{\beta}_{i,l}\widehat{m}_l(X_{i,j})\right\}^2 w(X_{i,j})$$

with respect to bandwidths. This is equivalent to minimizing:

$$\Xi_{AIC_1}=\sum_{i,j}\left\{Y_{i,j}-\sum_{l=0}^{L}\widehat{\beta}_{i,l}\widehat{m}_l(X_{i,j})\right\}^2 w(X_{i,j})\exp\left\{\frac{2L}{N}K_h(0)\int w(u)du\right\}$$

or computationally more easy criterion:

$$\Xi_{AIC_2}=\sum_{i,j}\left\{Y_{i,j}-\sum_{l=0}^{L}\widehat{\beta}_{i,l}\widehat{m}_l(X_{i,j})\right\}^2 \exp\left\{\frac{2L}{N}K_h(0)\frac{\int w(u)du}{\int w(u)\widehat{p}(u)du}\right\}.$$

Since the distribution of the data is very unequal the weight function w should give greater weight for the regions where data is sparse. One possible selection of w is $w(u)=\frac{1}{\widehat{p}(u)}$. Then the two criteria are:

$$\Xi_{AIC_1}=\sum_{i,j}\left\{Y_{i,j}-\sum_{l=0}^{L}\widehat{\beta}_{i,l}\widehat{m}_l(X_{i,j})\right\}^2 \widehat{p}(X_{i,j})\exp\left\{\frac{2L}{N}K_h(0)\int\frac{1}{\widehat{p}(u)du}\right\} \tag{3.21}$$

and

$$\Xi_{AIC_2}=\sum_{i,j}\left\{Y_{i,j}-\sum_{l=0}^{L}\widehat{\beta}_{i,l}\widehat{m}_l(X_{i,j})\right\}^2 \widehat{p}(X_{i,j})\exp\left\{\frac{2L}{N}K_h(0)\mu^{-1}\int\frac{1}{\widehat{p}(u)du}\right\}, \tag{3.22}$$

where μ is the measure of the design set.

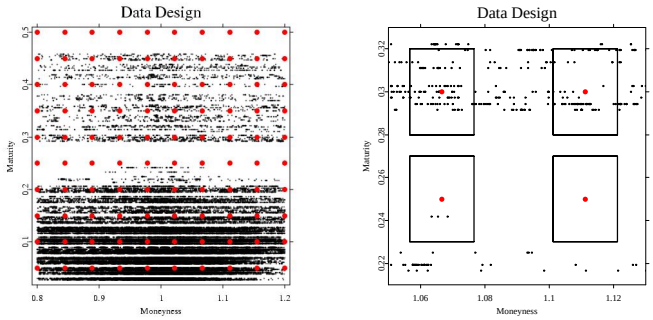


Figure 3.1: Left panel: pooled observation from January, 4th 1999 to March 8th, 1999. The large points are the hypothetical grid points on which the basis functions are evaluated. Right panel: the magnification of the left panel. The neighborhood of the points is marked with the rectangles.

3.5 Local bandwidths

In derivative market one can observe fairly many different types of option contracts. Each day one may trade options with several different time to maturities and many different strikes. However the number of possible strikes is much higher than the number of maturities, which results in the string structure. Moreover the contracts with smaller maturities are traded more intensively and there tend to exist more contracts for the smaller time to maturities for which the difference between two successive expiry days is one month (1M, 2M, 3M). For the next maturity range it increases to three months (6M, 9M, 12M).

Since the strings are moving in the maturity vs. moneyness plane towards expiry one needs to pool many days in order to fill the plane with observations. However due to an unequal distribution of data points one needs even more days to fill the range with bigger maturities than with smaller ones. Otherwise one faces gaps for some particular maturity range.

These gaps may obstruct the estimation procedure. If in any point u' the function $\widehat{p}(u') = 0$ in (3.3) then obviously matrix $B^{(r)}(u')$ in (3.8) contains only 0 and is singular. This means that one may not estimate successfully any value of the IVS in this point.

The problem with gaps is illustrated in Figure 3.1. Left panel presents pooled observation from January, 4th 1999 to March 8th, 1999. The large points are hypothetical

grid points. It is clearly visible that not all points are equally surrounded with the data. In the right panel magnification of the "problematic" grid points is displayed. If the particular grid point u' has no observations it the neighborhood than $\widehat{p}(u') = 0$ and $B^{(r)}(u')$ is singular.

The natural solution to this problem is increasing the bandwidths. However it may result in larger bias. Another possibility is to use the k-nearest neighbor estimator. In the range with many data, however, one takes into consideration only very few observations closest to the grid points. On the other hand in the range with few points the estimator is based on the observations far from the grid points. We propose different approach to cope with degenerated design. Instead of fixed bandwidths one may take local bandwidths, which vary according to the data density yielding smaller smoothing parameter in range with many data and bigger one where the data are sparse.

$$\widehat{p}_i(u) = \frac{1}{J_i}\sum_{j=1}^{J_i} K_{h(u)}(u - X_{i,j}), \tag{3.23}$$

$$\widehat{q}_i(u) = \frac{1}{J_i}\sum_{j=1}^{J_i} K_{h(u)}(u - X_{i,j})Y_{i,j}. \tag{3.24}$$

Our choice of the local bandwidths is motivated motivated by the approach of Gijbels and Mammen (1998). First choose fixed pilot bandwidths g. They minimize (3.22) or (3.21). Then plug-in local bandwidths according to density of the data:

$$h(u) = \left(\frac{\min \widehat{p}(u)}{\widehat{p}(u)} - \frac{\min \widehat{p}(u)}{\max \widehat{p}(u)} + 1\right)^{\delta} g \wedge g_{max} \tag{3.25}$$

where $\min \widehat{p}(u)$ and $\max \widehat{p}(u)$ are minimum and maximum values of $\widehat{p}(u)$ on the desire estimation grid. The bandwidths are smallest near the greatest density of the data and we believe that optimal bandwidths in this particular range are close to the bandwidths obtained in the pilot estimation. In the mode of the density of the data

$$\left(\frac{\min \widehat{p}(u)}{\widehat{p}(u)} - \frac{\min \widehat{p}(u)}{\max \widehat{p}(u)} + 1\right) = \left(\frac{\min \widehat{p}(u)}{\max \widehat{p}(u)} - \frac{\min \widehat{p}(u)}{\max \widehat{p}(u)} + 1\right) = 1$$

so the local bandwidths are equal to pilot bandwidths. The local bandwidths increase with the inverse of the data density. For the smallest density value they reach

$$\left(\frac{\min \widehat{p}(u)}{\min \widehat{p}(u)} - \frac{\min \widehat{p}(u)}{\max \widehat{p}(u)} + 1\right)^{\delta} g = \left(2 - \frac{\min \widehat{p}(u)}{\max \widehat{p}(u)}\right)^{\delta} g.$$

The parameter δ allows to control the maximum of the smoothing parameters. It could be also controlled by g_{max} where the bound on the bandwidths is imposed.

3.6 Initial parameters selection

The problem of gaps in the data cannot only be handled with the size of the bandwidths. Of course it is obligatory that $\widehat{p}_i(u)$ needs to be non-zero for at least one i. However this is not a sufficient condition to ensure non singularity of the matrix $B^{(r)}(u')$. The initial estimates of $\widehat{\beta}_i^{(0)}$ play also an important role.

In Fengler et al. (2005) a piecewise constant initial time series are proposed. The subintervals $I_1, ..., I_L$ are pairwise disjoint subsets of $\{1, ..., I\}$ and $\bigcup_{l=1}^{L} I_l$ is a strict subset of $\{1, ..., I\}$. The initial estimates are now defined by $\widehat{\beta}_{i,l}^{(0)} = 1$ if $i \in I_l$ and $\widehat{\beta}_{i,l}^{(0)} = 0$ if $i \notin I_l$. To complete the setting $\widehat{\beta}_{i,0}^{(0)} = 1$ for each i.

However this kind of setting requires even more data to obtain the final estimates. For each subset I_l there needs to exist at least one day i such that $\widehat{p}_i(u') \neq 0$, otherwise the row of zeros in (3.8) appears. The smaller is the length of I_l intervals the bigger bandwidths need to be taken. This deficiency can be removed by taking a random initial time series. Then $\widehat{p}_i(u)$ needs to be non-zero for one i in $\{1, ..., I\}$ and it is no longer necessary that $\widehat{p}_i(u)$ is non-zero for one i in each I_l.

4 Implementation Issues

The main aim of modelling using the DSFM is to approximate the IVS with low-dimensional representation. One faces the problem with high complexity of the data structure and has to consider great quantity of observations. On the particular day even on the daily level there are approximately 80 observations and on the intra-day level the number of observations rises to over 2000 per day. Additionally in order to reflect the dynamics of the whole IVS the model needs to be estimated on the sufficiently large time interval. This fact causes that the number of observations increase even more.

In order to cope efficiently with huge amount of the data the optimal implementation is important. The speed of the algorithm may highly dependent on the proper order of calculations. The double calculations of the same quantities should be avoided, because it may significantly slow down the estimation procedure.

In this chapter we consider some numerical issues, which need to be taken into account for efficient DSFM's implementation. In Section 4.1 we discuss numerical methods used in the estimation algorithm and typical numerical problems for DSFM implementation is considered in Section 4.2. Section 4.3 focusses on the XploRe implementation, which is done as a part of this thesis, and Section 4.4 presents the efficiency study of the algorithm.

Throughout this chapter we keep notation for I as a number of observed days and length of the $\widehat{\beta}_l$ time series. We also set M_u to a total number of grid points, on which $\widehat{m}_l$ functions are evaluated.

4.1 Numerical Algorithms

In this section we present the numerical algorithms which are used for the estimation of the DSFM. The methods essential in successful estimation are briefly discussed. For the detailed description we refer to Press et al. (1992).

4.1.1 LU Decomposition

In (3.10) and (3.15) one has to solve the system of linear equations. Obviously for each grid point u' in (3.10) the different linear system appears. It means each system has to be solved separately. Similarly for each time point i one needs to solve (3.15) in order to obtain estimates for $\widehat{\beta}_{i,l}$. However since the DSFM is applied also for the dimension reduction each particular linear system is typically low-dimensional ($L+1 \times L+1$ for the (3.10) and $L \times L$ for (3.15)). In each iteration update one needs to solve $I + M_u$ linear systems and this is the numerical challenge.

Among many methods for solving the linear system of equations LU decomposition gives fast and accurate approximation of the solution. The matrix $\mathbf{A}$ is given as product of two matrices $\mathbf{L}$ and $\mathbf{U}$

$$\mathbf{A} = \mathbf{LU},$$

where $\mathbf{L}$ is lower triangular (has nonzero elements only on the diagonal and below) and $\mathbf{U}$ is upper triangular (has nonzero elements only on the diagonal and above). The LU decomposition by the elements for the $n \times n$ matrix is given by:

$$\begin{pmatrix} l_{11} & 0 & \dots & 0 \\ l_{21} & l_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ l_{n1} & l_{n2} & \dots & l_{nn} \end{pmatrix} \cdot \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1n} \\ 0 & u_{22} & \dots & u_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & u_{nn} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}.$$

After obtaining the decomposition the solution of the linear set is straightforward. One may rearrange it in following way:

$$\mathbf{Ax} = \mathbf{LUx} = \mathbf{L}(\mathbf{Ux}) = \mathbf{Ly} = \mathbf{b}$$

where $\mathbf{x}$ is unknown and $\mathbf{b}$ is known vector. First

$$\mathbf{Ly} = \mathbf{b} \tag{4.1}$$

is solved and afterwards

$$\mathbf{Ux} = \mathbf{y}. \tag{4.2}$$

Since the matrices $\mathbf{L}$ and $\mathbf{U}$ are triangular the solutions of (4.1) and (4.2) can be easily obtained by forward and backward substitution.

For the estimation of the DSFM the matrix $\mathbf{A}$ is $M^{(r)}(i)$ or $B^{(r)}(u)$. The vector $\mathbf{b}$ is $S^{(r)}(i)$ or $Q^{(r)}(u)$ and the unknown vector $\mathbf{x}$ is either the estimates of the loadings in time i - $\widehat{\beta}_i^{(r)}$ - or the estimates of the basis function in particular point u' - $\widehat{m}^{(r)}(u')$.

The remaining challenge is to calculate the matrices $\mathbf{L}$ and $\mathbf{U}$. Each element of the matrix $\mathbf{A} = \mathbf{a_{kj}}$ is given as a sum of elements of $\mathbf{L}$ and $\mathbf{U}$

$$l_{k1}u_{1j} + ... + l_{kn}u_{nj} = a_{kj}.$$

This results in n^2 equations (for each a_{ij}) with $n^2 + n$ unknowns (nonzero elements of both triangular matrices). The greater number of unknowns that equations suggest that the n of unknowns can be specified arbitrarily. The natural choice would be as in Press et al. (1992) to specify the diagonal of $\mathbf{L}$ or $\mathbf{U}$. Let from now on $l_{jj} = 1$ for $j = 1, ..., n$.

The n equations can be now written as follows:

$$\begin{aligned}
u_{11} &= a_{11} \\
l_{21}u_{11} &= a_{21} \\
l_{31}u_{11} &= a_{31} \\
\vdots &\quad \vdots \\
l_{n1}u_{11} &= a_{n1} \\
u_{12} &= a_{12} \\
l_{21}u_{12} + u_{22} &= a_{21} \\
l_{31}u_{12} + l_{32}u_{22} &= a_{32} \\
\vdots &\quad \vdots
\end{aligned}$$

Note that for each equation on the left side there exist only one number that did not appeared in previous equations. That means that this system can be solved sequentially. First u_{11} is set to a_{11} then l_{21} is calculated and so on.

Since the diagonal of L is set to 1 only n values needs to be kept in memory. The LU decomposition can be remembered as one matrix:

$$\begin{pmatrix}
u_{11} & u_{12} & \dots & u_{1n} \\
l_{21} & u_{22} & \dots & u_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
l_{n1} & l_{n2} & \dots & u_{nn}
\end{pmatrix}$$

4.1.2 Eigensystems

In (3.17) and (3.18) the calculations of inverse of the symmetric matrix Γ and also calculations of $\Gamma^{-\frac{1}{2}}$ are required. The power of any symmetric matrix $\mathbf{A}$ can be obtained from Jordan decomposition:

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^\top,$$

where Λ is a diagonal matrix consisting eigenvalues of $\mathbf{A}$. $\mathbf{Q}$ is an orthogonal matrix of eigenvectors of $\mathbf{A}$. The power of the matrix is given then:

$$\mathbf{A}^\alpha = \mathbf{Q}\mathbf{\Lambda}^\alpha\mathbf{Q}^\top,$$

where power of diagonal matrix Λ^α is defined as a power of all matrix's elements.

In order to obtain the Jordan decomposition we propose the Jacobi method. We follow the recommendations of Press et al. (1992) to use this method for *matrices of moderate order*. More complex but faster methods like QR can be used too. Since the Jacobi algorithm is supposed to give efficiently good results for matrices of order smaller than 10 and the matrix Γ has order L, which is typically small, we believe that the choice of this method does not affect significantly the speed of the whole estimation.

The Jacobi method is based on sequence of orthogonal transformations with rotation matrix $\mathbf{P_{pq}}$ of the form

$$\mathbf{P_{pq}} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & c & \cdots & s & & \\ & & \vdots & 1 & \vdots & & \\ & & -s & \cdots & c & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}$$

where s and c denote sine and cosine of the rotation angle. All elements off-diagonal elements are zero except s and $-s$. All elements on the diagonal are one except c in row/column p and q. The diagonal matrix Λ is obtained as transformation with the sequence of rotation matrices:

$$\Lambda = \cdots\mathbf{P_2}^\top\mathbf{P_1}^\top\mathbf{A}\mathbf{P_1}\mathbf{P_2}\cdots = \mathbf{V}^\top\mathbf{A}\mathbf{V}$$

where $\mathbf{V} = \mathbf{P_1 P_2} \cdot \cdot$ is a sequence of some rotation matrices. Since $\mathbf{V}$ is orthogonal as a product of orthogonal matrices, the Jordan decomposition is then given by:

$$\mathbf{A} = \mathbf{V \Lambda V}^\top.$$

Each transformation $\mathbf{P}_{\mathbf{pq}}^\top \mathbf{A}$ changes only row p and q. Similarly $\mathbf{A P_{pq}}$ changes only columns p and q.

$$\mathbf{A}' = \begin{pmatrix} & \cdots & a'_{1p} & \dots & a'_{1q} & \cdots & \\ \vdots & & \vdots & & \vdots & & \vdots \\ a'_{p1} & \cdots & a'_{pp} & \dots & a'_{pq} & \cdots & a'_{pn} \\ \vdots & & \vdots & & \vdots & & \vdots \\ a'_{q1} & \cdots & a'_{qp} & \dots & a'_{qq} & \cdots & a'_{qn} \\ \vdots & & \vdots & & \vdots & & \vdots \\ & \cdots & a'_{np} & \dots & a'_{nq} & \cdots & \end{pmatrix}$$

The sequence of rotation matrices is arranged such that new matrix

$$\mathbf{A}' = \mathbf{P}_{\mathbf{pq}}^\top \mathbf{A P_{pq}}$$

has zeros both on p's column q's row and on q's column p's row ($a'_{pq} = a'_{qp}$). In order to calculate other elements of $\mathbf{A}'$ first set $\theta = \frac{a_{qq} - a_{pp}}{2a_{pq}}$ and $t = \frac{sgn(\theta)}{|\theta| + \sqrt{\theta^2 + 1}}$. The cosine and sine are respectively:

$$\begin{aligned} c &= \frac{1}{\sqrt{t^2 + 1}} \\ s &= tc. \end{aligned}$$

Let

$$\tau = \frac{s}{1 + c}$$

then other elements of $\mathbf{A}'$ which has to be changed are given by:

$$\begin{aligned} a'_{pp} &= a_{pp} - t a_{pq} \\ a'_{qq} &= a_{qq} + t a_{pq} \\ a'_{rp} &= a_{rp} - s(a_{rp} + \tau a_{rp}) \\ a'_{rq} &= a_{rq} + s(a_{rp} - \tau a_{rq}) \end{aligned}$$

where $r = 1, ..., n$ and $r \neq p, q$. Note that if element a_{pq} is annihilated by orthogonal transformation it does not necessarily stays zero after different orthogonal transformation. One has to apply rotation to all elements and then repeat whole procedure. The set of rotations for each element of matrix $\mathbf{A}$ is called *sweep*. The sweeps need to be repeated couple of times until the off-diagonal matrix are sufficiently small (numerical zeros). Typical matrices require 6 to 10 sweeps. The order of annihilation the elements in the sweep is not really important. We propose after Press et al. (1992) to proceed down the rows $\mathbf{P_{12}}, \mathbf{P_{13}}, \mathbf{P_{14}}$ then $\mathbf{P_{23}}, \mathbf{P_{24}}, ...$ etc.

4.2 Numerical Difficulties of the DSFM

In Section 4.1 we consider algorithms for solving numerically problems which arise in DSFM estimation. The other problems which appears are elementary calculations. However DSFM is supposed to perform the estimation from high dimensional intra-day data and these elementary calculations has to be arrange optimally in order to save memory and increase the speed. In this section we consider some numerical difficulties, which appear in DSFM's estimation.

As a numerical result of the estimation L time series $(\widehat{\beta}_{i,l})$ and $L+1$ functions $(\widehat{m}_l)$ are obtained. The basis factor functions can be represented on the finite grid only, which obviously may not cover the whole desired estimation space. This grid has to be set in the range of interest at the very beginning of the estimation procedure. One has to remember $L+1$ functions on M_u grid points in each step of the algorithm. Obviously, L time series of length I for $\widehat{\beta}_{i,l}$ has to kept constantly in memory as well.

All data from the option market, denoted as $X_{i,j}$ and $Y_{i,j}$ play a role only in calculating $\widehat{p}_i(u)$ in (3.3) and $\widehat{q}_i(u)$ in (3.4). After calculating these quantities the memory for $X_{i,j}$ and $Y_{i,j}$ can be freed. Obviously we have to remember $2 \times I \times M_u$ numbers but mostly it is significantly smaller than the amount of memory necessary for keeping the whole data. Once calculated $\widehat{p}_i(u)$ and $\widehat{q}_i(u)$ could be used in estimation for different L. They can be also reused for estimation for another time period, which includes some part from the old one. That is why we believe that "sliding window" type of analysis is feasible for DSFM. One has to simply remove from the memory $\widehat{p}_1(u)$ $(\widehat{q}_1(u))$ and replace with $\widehat{p}_{I+1}(u)$ $(\widehat{q}_{I+1}(u))$, while the rest stays same. Unfortunately changing bandwidths h requires recalculation of all $\widehat{p}_i(u)$ and $\widehat{q}_i(u)$.

Except of keeping in memory $\widehat{p}_i(u)$ and $\widehat{q}_i(u)$ one may also remember $J_i\widehat{p}_i(u)$ and $J_i\widehat{q}_i(u)$. All this quantities play important role in computing $B^{(r)}(u)$, $Q^{(r)}(u)$, $M^{(r)}(i)$ and $S^{(r)}(i)$ in each iteration step. Because this quantities are computed many times, $B^{(r)}(u)$, $Q^{(r)}(u)$ for each grid point, $M^{(r)}(i)$ and $S^{(r)}(i)$ for each time point, it is essential that they are obtained fast. It is also possible to keep in memory $\widehat{\beta}_{i,l'}^{(r-1)}\widehat{\beta}_{i,l}^{(r-1)}$

for $0 \leq l, l' \leq L$ and $i = 1, ..., I$ since this product appears for each grid point in $B^{(r)}(u)$. Similarly the calculations of $\widehat{m}_{l'}(u)\widehat{m}_l(u)$ for $0 \leq l, l' \leq L$ and each grid point can be optimized. In (3.13) and (3.14) the integrals $\int \widehat{p}_i(u)\widehat{m}_{l'}(u)\widehat{m}_l(u)\, du$ have to be calculated. Numerically they are approximated by appropriate sum of function values given on the grid.

Since the matrices $B^{(r)}(u)$ and $M^{(r)}(i)$ are symmetric it is not a good idea to calculate the whole matrix. It is enough to compute lower triangular (upper triangular) part of the matrix and simply copy other elements.

Finally we have to state that except keeping in memory the actual estimates of $\widehat{m}_l^{(r)}$ and $\widehat{\beta}_{i,l}^{(r)}$ the estimates from one previous step of iteration $\widehat{m}_l^{(r-1)}$ and $\widehat{\beta}_{i,l}^{(r-1)}$ need to be remembered in order to stop the iteration as it is given in (3.16).

4.3 XploRe Implementation

This section considers XploRe implementation of the DSFM. New quantlets `DSFM` and `DSFM4IV` were implemented for efficient estimation. Although they are part of statistical programming language XploRe but the core of the algorithm is done as external DLL library in order to improve the speed. This solution allows as well to import the algorithm to any environment which supply DLL communication.

```
{beta,mhat} = DSFM(x,L,h,u,rmax,L2tol,startbeta)

{beta,mhat,fitted} = DSFM4IV(x,L,h,u,rmax,L2tol,startbeta)
```

Estimates a dynamic semiparametric factor model from the form:$y_t = m_0(u) + \sum_{l=1}^{L} \beta_{i,l} m_l(u)$,, where $\widehat{m}_0$ to $\widehat{m}_L$ are 2-dimensional invariant basis functions on the grid u and β_1 to β_L are scalar weights depending on time T. After estimation, the functions m are orthogonalized under the empirical norm of the observed data points and ordered according to the maximal variance of β.

As a result of the `DSFM` quantlet two matrices `beta` and `mhat` are obtained. `mhat` is $M_u \times (L+3)$ matrix and contains the grid points in first and second column and values

of all functions given on the grid in other columns. $\widehat{m}_0$ is then on the third column $\widehat{m}_1$ fourth and so on. `beta` is $I \times (L+2)$ matrix and contains dates in the first column $\widehat{\beta}_0$ in the second one, $\widehat{\beta}_1$ in the third one etc. Since $\widehat{\beta}_0$ always contains only ones it does not need to be remembered. However this form is kept in this way to simplify the calculation of the IVS, which in XploRe code is given simply by:

```
IVS=mhat[,1:2] ~ sum(beta[1,2:cols(beta)].*mhat[,3:cols(mhat)],2)
```

for the first day of the sample.

The quantlet has seven different input parameters. The data are given in `x`, which is four column matrix. First two columns contains the coordinates of the grid $(X_{i,j})$, the third column observed function value $(Y_{i,j})$ and the last on indictor for the day - typically date. `L` is a scalar and denotes the number of dynamic basis functions or factor loadings. The parameter `h` gives the bandwidths for the kernel function. It can be either 2×1 vector (for global bandwidths) or $M_u \times 2$ vector (for local bandwidths). The local bandwidths are applied only if the dimension of the `h` is the same as dimension of the `u`, namely $M_u \times 2$. `u` is a regular grid on which estimation is performed.

Except the parameters mentioned above the user may also specify the optional parameters. `rmax` stays for the maximum number of iterations. The default value is set to 301. `L2tol` is minimal L_2 tolerance of the estimation criterion denoted by ϵ in (3.16). The default value is 10^{-5}. In `startbeta` the user may set starting parameters for initial factor loadings. If it is not given locally constant starting time series proposed in Fengler et al. (2005) are used.

The second quantlet `DSFM4IV` is fairly similar to `DSFM`. It is an extension for IV analysis. We believe that DFSM can be applied also in other fields so we supply two quantlets. Except the estimates `DSFM4IV` also gives a table with summary of the data, model parameters, explained variance, Ξ_{AIC_1} and Ξ_{AIC_2} criteria. As an output a matrix `fitted` is given additionally. Its six columns contain respectively: date, the number of observation in each day, mean of Ξ_{AIC_1} criterion, variance of Ξ_{AIC_1} criterion, mean of Ξ_{AIC_2} criterion and variance of Ξ_{AIC_2} criterion.

4.4 Efficiency of the Algorithm

The estimation procedure described in Chapter 3 consists of three consequent parts. First the values of $\widehat{p}_i(u)$ and $\widehat{q}_i(u)$ have to be calculated. Then the iteration and orthogonalization is proceeded. Finally some characteristic of the fit like (3.20), (3.21) or (3.21) may be additionally computed.

Part	Time(s)	Percentage
$\widehat{p}_i$ and $\widehat{q}_i$ calculation	271.652	40.613%
Iteration	12.250	1.832%
Orthogonalization and ordering	0.031	0.005 %
Model characteristics calculation	384.813	57.550 %
Joint	668.656	100%

Table 4.1: Time in seconds which is needed for the inner parts of the estimation algorithm.

Out of these main parts the iteration and the orthogonalization are not so computationally intensive like the remaining ones. These is due to the fact then one does not need to operate on the data itself but only on some already pre-computed characteristics of the data. In order to make it more sound we present in Table 4.1 the time in seconds which is necessary for the estimation of the each part for the data set used in Chapter 5.

The actual time in seconds is highly dependent on the used CPU but the percentage bring some important information. It is not surprising that the iteration and orthogonalization parts together takes not more than 2% of the joint computational time. This findings approve that under some specific limitations one may efficiently do "sliding window" type of analysis. However the bandwidths should be then fixed for whole time interval and one have to resign from calculating model characteristics.

5 Applications

The application ability of any model depends on its complexity and efficient estimation algorithm. The algorithm, numerical issues and estimation procedure is discussed in previous chapters. In this chapter we discuss the possible application of the DSFM. We limit ourself to application to the options on DAX but application to options on other indexes or in different fields are also possible.

First we present the data set which is used. Then in Section 5.2 we discuss the estimation results, the selection of the bandwidths, choice of the number of the basis functions and the starting time series. Section 5.3 presents the model performance on the simulated data. Finally we present hedging application of the DSFM.

The aim of this chapter however is not to discuss all the details of DSFM applications. We give only short overview in which areas the model can be used. We believe that the applications, which are presented (and other applications, which are not discussed here), can be studied more intensively due to the efficient implementation of the model.

5.1 Data

For our analysis we employ tick statistics on DAX index options from January 1999 to February 2003. By inverting the BS formula one easily obtains IV. We regard as outliers observations with IV bigger that 0.8 and smaller than 0.04. We also remove observations with maturity less than 10 day since their behavior in this range is irregular due to expiry effect. The data set contains 1054 days of observations and approximately 2.8 millions of contracts i.e. around 2700 per day. We have both call and put options. Violation of the call-put parity is removed by Hafner-Wallmeier correction. For the details we refer to Hafner and Wallmeier (2001).

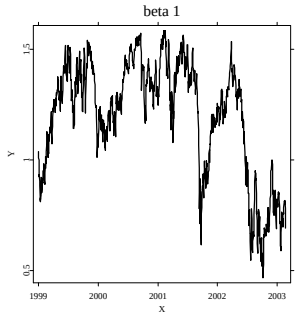


Figure 5.1: Time series of loadings $\widehat{\beta}_1$.

5.2 Estimation results

We apply the algorithm on an equidistant grid covering moneyness $\kappa \in [0.8, 1.2]$ and time to maturity measured in years $\tau \in [0.05, 1.00]$. In each direction our grid consists of 25 points. We apply local bandwidths and the number of dynamic factor functions L we set to 3. As the starting values of $\widehat{\beta}^{(0)}$ we take a piecewise constant series on disjoint time intervals.

Figures (5.1)-(5.2) present the estimated factors loading $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$. The magnitude and variance of the $\widehat{\beta}_1$ are much higher than for the other two time series, which suggests that the first basis function has the biggest explanatory power of the IVS variation. This is actually not surprising since the basis functions were ordered with respect to the biggest variance of loading factors.

Figure (5.3) displays the estimated basis functions $\widehat{m}_0$ - $\widehat{m}_3$. We find similar interpretations of the factors as in Fengler et al. (2005) or Cont and da Fonseca (2002). The first dynamic factor $\widehat{m}_1$ is relatively flat on almost the whole range and negative on all grid points. It reflects the up and down shifts of the entire log-IVS. For the small maturities a strong curvature can be seen. It corresponds to the empirical fact that near the expiry the smile effect becomes stronger. The second function is positive for the small maturities and negative for the bigger maturities. The positive $\widehat{\beta}_2$ increases short term maturities IVs and simultaneously decreases the long term ones. The negative $\widehat{\beta}_2$ causes the opposite effect. This function provides term structure changes of the IVS. The last function $\widehat{m}_3$ reveals a strong slope in the moneyness direction changing from positive to negative near at-the-money. It reflects changes of the moneyness slope and smile curvature.

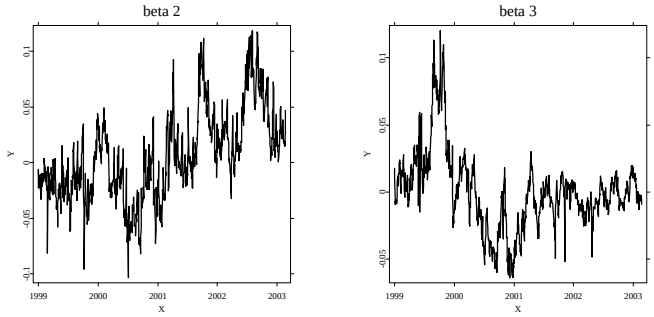


Figure 5.2: Time series of loadings $\widehat{\beta}_2$ and $\widehat{\beta}_3$.

The DSFM's fit to the data on January, 4th 1999 is presented in Figure (5.4). For the comparison also the fit obtained with Nadaraya-Watson (NW) kernel estimator is displayed. The bandwidths for NW estimator are $h_1 = 0.04$ in moneyness direction and $h_2 = 0.06$ in time to maturity direction. They correspond to the local bandwidths used in DSFM. The selection of the bandwidths is described in Section 5.2.1. On this particular day the DSFM fit is not better than NW fit. It is due to the fact that DSFM reflects the dynamics of IVS and the smoothing is done not only in space but also in time. DSFM yields always the estimate on the whole considered space since the IVS is a weighted sum of the factors. However it is not always the case for NW estimator. It may not give reasonable estimates in places with no observations. The upper panels of Figure 5.4 illustrate this problem. Although DSFM estimates the smile for the whole range of moneyness $[0.8, 1.2]$ the NW yields no estimates near $\kappa = 1.2$. To make it even more sound we present the estimate of the whole IVS in Figure 5.5. The left panel shows the estimate of IVS with NW estimator. The surface is well defined only near the strings while the surface obtained by DSFM covers the whole considered space. In order to remove the NW deficiencies one needs to increase the bandwidths which results in larger bias as it was presented in Fengler et al. (2005).

5.2.1 Bandwidths selection

For the data-driven choice of bandwidths we follow the algorithm described in Section 3.5. First we apply the estimation procedure on several possible bandwidths. The result of the pilot estimation is presented in Table 5.1.

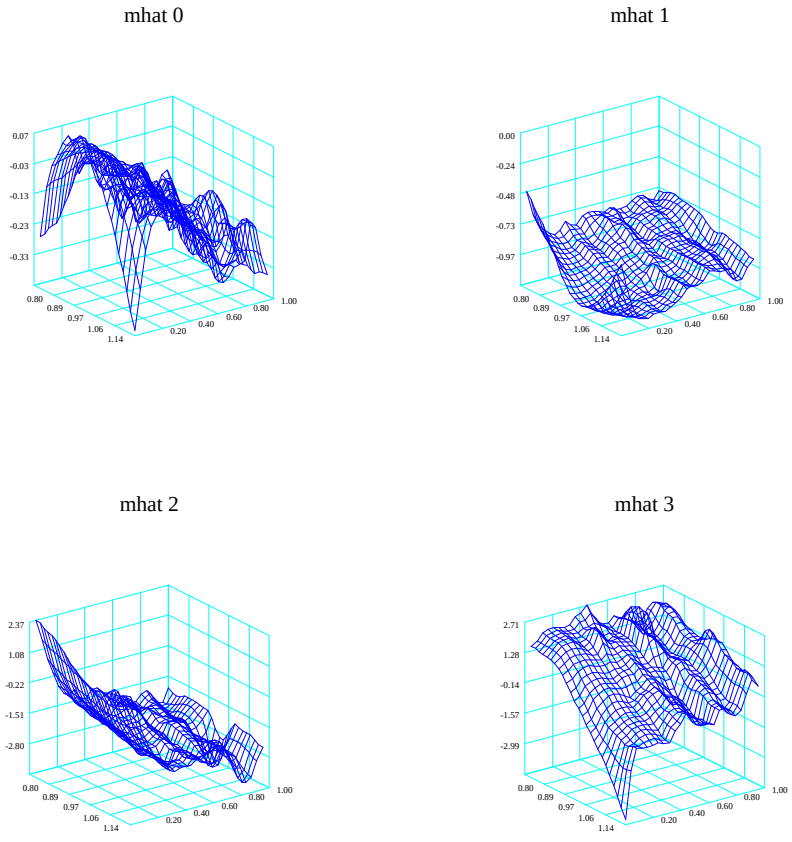


Figure 5.3: Invariant basis function $\widehat{m}_0$ and dynamic basis functions $\widehat{m}_1$, $\widehat{m}_2$ and $\widehat{m}_3$.

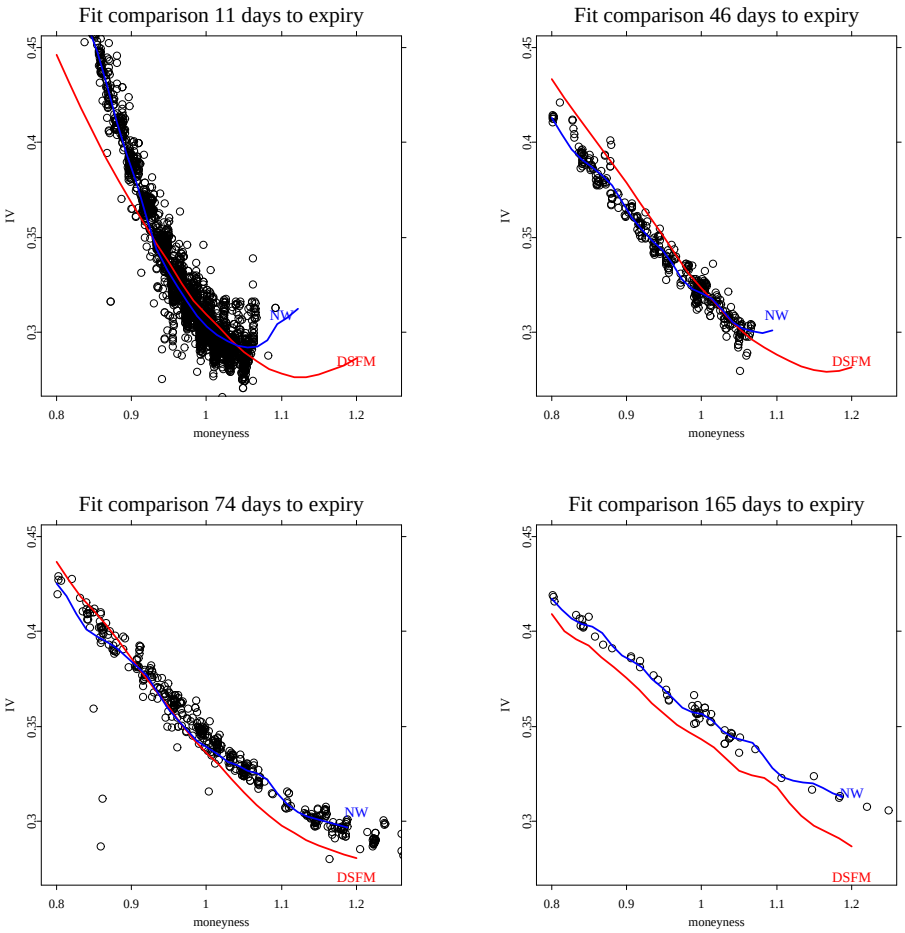


Figure 5.4: Comparison of the fits obtained with DSFM and Nadaraya-Watson estimator with $h_1 = 0.04$ and $h_2 = 0.06$ on January, 4th 1999.

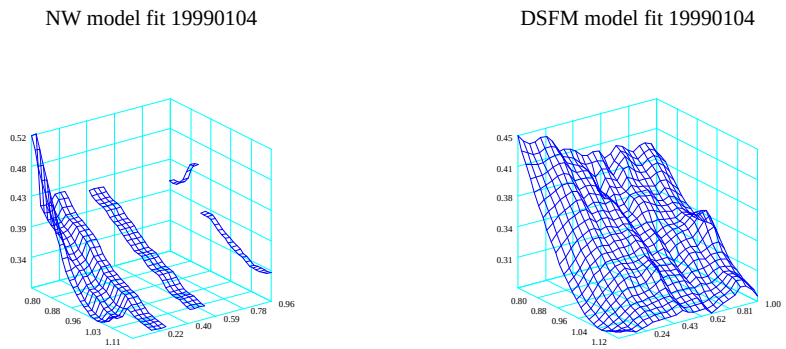


Figure 5.5: Left panel: NW IVS fit on January, 4th 1999. Right panel: DSFM fit on January, 4th 1999

h_1	h_2	AIC1	AIC2	$RV(3)$
0.01	0.04	0.001893	0.001715	0.982043
0.01	0.06	0.001860	0.001671	0.982082
0.01	0.08	0.001826	0.001680	0.981770
0.01	0.10	0.001799	0.001707	0.981348
0.02	0.02	0.002300	0.002350	0.975581
0.02	0.04	0.001841	0.001641	0.982201
0.02	0.06	0.001822	0.001623	0.982193
0.02	0.08	0.001798	0.001644	0.981862
0.02	0.10	0.001777	0.001678	0.981430
0.03	0.02	0.002254	0.002263	0.975844
0.03	0.04	0.001827	0.001620	0.982234
0.03	0.06	0.001812	0.001611	0.982205
0.03	0.08	0.001790	0.001635	0.981867
0.03	0.10	0.001771	0.001671	0.981430
0.04	0.02	0.002216	0.002199	0.976221
0.04	0.04	0.001818	0.001615	0.982185
0.04	**0.06**	0.001805	0.001611	0.982134
0.04	0.08	0.001785	0.001638	0.981788
0.04	0.10	0.001768	0.001675	0.981345
0.05	0.02	0.002188	0.002163	0.976436
0.05	0.04	0.001814	0.001627	0.981998
0.05	0.06	0.001802	0.001627	0.981920
0.05	0.08	0.001784	0.001655	0.981568
0.05	0.10	0.001767	0.001693	0.981120
0.06	0.02	0.002167	0.002152	0.976444
0.06	0.04	0.001814	0.001659	0.981602
0.06	0.06	0.001803	0.001662	0.981502
0.06	0.08	0.001785	0.001692	0.981142
0.06	0.10	0.001770	0.001730	0.980692

Table 5.1: The pilot bandwidths selection via AIC for the different choices of $h = (h_1, h_2)^\top$. h_1 refers to moneyness and h_2 to time to maturity.

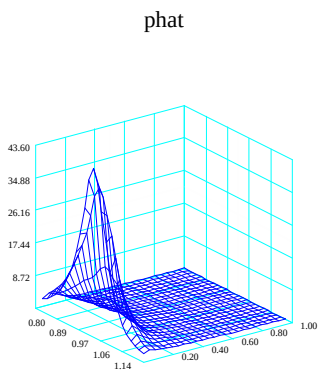


Figure 5.6: The overall density of observations $\widehat{p}(u)$.

The pilot bandwidths are chosen to $h_1 = 0.04, h_2 = 0.06$ according to Ξ_{AIC_2}. However in the neighborhood both criteria Ξ_{AIC_1} and Ξ_{AIC_2} become very flat. In order to obtain the local bandwidths we apply (3.25) with arbitrary chosen $\delta = 1$ and g_{max} set to $\frac{1}{3}$ of the estimation space. The overall density $\widehat{p}$ is presented in Figure (5.6). It clearly visible that the biggest density of the observations is at-the-money for small maturities.

The chosen bandwidths are presented in Figure (5.7) as functions of u. They are both relatively flat for the small maturities and increase rapidly for the bigger maturities. The bandwidths do not increase so fast in maturity for the moneyness close to 1. It reflects the fact that near at-the-money one may observe more trades than in other ranges.

5.2.2 Model selection

We model the dynamics of the IVS with $L = 3$ dynamic functions. The motivation of our choice is presented in Table 5.2. We have recalculated the model with the same bandwidths and starting time series for $1, 2, 3$ and 4 dynamic basis functions. Adding 4-th function does not have significant influence on the explained variance. For this reason we limit the model to $L = 3$ dynamic basis functions. Three functions are

L	$1 - RV(L)$	ΔRV
1	0.9638	
2	0.9739	0.0101
3	0.9822	0.0083
4	0.9830	0.0008

Table 5.2: Explained variance for the model size.

also considered in Fengler et al. (2005) or in other than DSFM models like Cont and da Fonseca (2002). Note that the model with only one dynamic basis function can explain 0.96% of the variance.

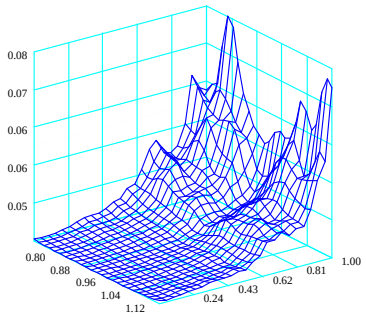
5.2.3 Initial parameter dependence

Apart from the bandwidth selection the choice of the initial starting parameters $\widehat{\beta}^{(0)}$ is another estimation issue. We have recalculated the estimates for different starting values. Denote by PC_1, PC_2, PC_3 the different settings of piecewise constant starting values as described in Section 3.6. Denote also by BM_1, BM_2, BM_3 settings where the algorithm starts from a Brownian Motion, WN_1, WN_2, WN_3 from a white noise and AR_1, AR_2, AR_3 from $AR(1)$ process. For each of the 12 settings we have obtained different estimates of basis functions and weights. The correlation matrices between obtained $\widehat{\beta}_1, \widehat{\beta}_2, \widehat{\beta}_3$ in each setting is given in Table 5.3.

All elements of the matrices are either 1 or -1, which means that the same solution is indicated for each of the starting $\widehat{\beta}^{(0)}$ since the weights are perfectly correlated. Of course if the correlation of the time series estimates is -1 the same factors are considered because they are identifiable only up to sign.

Table 5.3 suggest that starting time series do not have significant influence on the estimation results. However it is not always the case. Two different solution depending on the initial estimates were indicated in Borak et al. (2005), but smaller bandwidths for most of the data were used. The stability of the estimates with respect to starting time series seems to be bigger for bigger bandwidths. The values of the estimated basis functions with greater bandwidths depend then more on observations than on initial time series.

local bandwidths in moneyness



local bandwidths in maturity

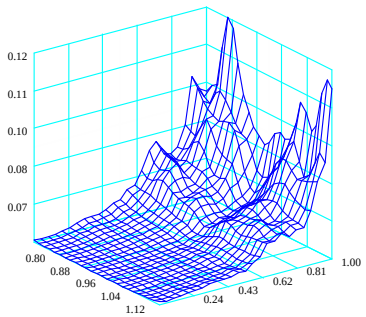


Figure 5.7: Local bandwidths used for the model estimation as functions of u.

$$
\begin{array}{c}
\\
PC_1 \\ PC_2 \\ PC_3 \\ BM_1 \\ BM_2 \\ BM_3 \\ WN_1 \\ WN_2 \\ WN_3 \\ AR_1 \\ AR_2 \\ AR_3
\end{array}
\begin{pmatrix}
1 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\
 & 1 & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\
 & & 1 & -1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\
 & & & 1 & -1 & -1 & -1 & -1 & 1 & 1 & -1 & -1 \\
 & & & & 1 & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\
 & & & & & 1 & 1 & 1 & -1 & -1 & 1 & 1 \\
 & & & & & & 1 & 1 & -1 & -1 & 1 & 1 \\
 & & & & & & & 1 & -1 & -1 & 1 & 1 \\
 & & & & & & & & 1 & 1 & -1 & -1 \\
 & & & & & & & & & 1 & -1 & -1 \\
 & & & & & & & & & & 1 & 1 \\
 & & & & & & & & & & & 1
\end{pmatrix}
$$

$$
\begin{array}{c}
\\
PC_1 \\ PC_2 \\ PC_3 \\ BM_1 \\ BM_2 \\ BM_3 \\ WN_1 \\ WN_2 \\ WN_3 \\ AR_1 \\ AR_2 \\ AR_3
\end{array}
\begin{pmatrix}
1 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\
 & 1 & 1 & -1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\
 & & 1 & -1 & -1 & -1 & -1 & 1 & -1 & 1 & 1 & 1 \\
 & & & 1 & 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\
 & & & & 1 & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\
 & & & & & 1 & 1 & -1 & 1 & -1 & -1 & -1 \\
 & & & & & & 1 & 1 & 1 & -1 & -1 & -1 \\
 & & & & & & & 1 & -1 & 1 & 1 & 1 \\
 & & & & & & & & 1 & -1 & -1 & -1 \\
 & & & & & & & & & 1 & 1 & 1 \\
 & & & & & & & & & & 1 & 1 \\
 & & & & & & & & & & & 1
\end{pmatrix}
$$

$$
\begin{array}{c}
\\
PC_1 \\ PC_2 \\ PC_3 \\ BM_1 \\ BM_2 \\ BM_3 \\ WN_1 \\ WN_2 \\ WN_3 \\ AR_1 \\ AR_2 \\ AR_3
\end{array}
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 \\
 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 \\
 & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 \\
 & & & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 \\
 & & & & 1 & 1 & 1 & 1 & 1 & 1 & -1 & 1 \\
 & & & & & 1 & -1 & -1 & -1 & -1 & 1 & -1 \\
 & & & & & & 1 & 1 & 1 & 1 & -1 & 1 \\
 & & & & & & & 1 & 1 & 1 & -1 & 1 \\
 & & & & & & & & 1 & 1 & -1 & 1 \\
 & & & & & & & & & 1 & -1 & 1 \\
 & & & & & & & & & & 1 & -1 \\
 & & & & & & & & & & & 1
\end{pmatrix}
$$

Table 5.3: The correlation matrices of factor loadings $\widehat{\beta}_i$ obtained for different initial starting time series $\widehat{\beta}^{(0)}$.

The importance of the random $\widehat{\beta}^{(0)}$ is seen for the smaller data sets. For one year of observations the estimation procedure of DSFM does not yield any results for bandwidths $h_1 = 0.04, h_2 = 0.06$ and local constant initial time series. The matrix $B^{(r)}(u')$ is simply singular for some u'. The enlarging bandwidths does not solve the problem, because local constant weights also lead to a singular matrix. However keeping bandwidths $h_1 = 0.04, h_2 = 0.06$ and employing white noise for $\widehat{\beta}^{(0)}$ provide successfully estimates.

5.3 Simulated example

In Section 5.2 we apply the DSFM to the option data. In this section we apply the model to artificial generated data in order to check how well the estimation procedure can reproduce predefined structure. By applying the algorithm to simulated data one can compare the obtained estimates of the factor loadings and factor functions with real ones, from which the data are generated.

The data are simulated on the grid containing 625 points, 25 in both moneyness and time to maturity directions. The estimation space cover $\kappa \in [0.8, 1.2]$ and $\tau \in [0.0, 1.0]$. The time period is limited to 200 days and in each day 1000 observation are simulated. For the simplicity we do not impose the classical IV string structure but fill the space uniformly with data.

First for each $j = 1, ..., 200$ and $i = 1, ..., 1000$ the uniformly distributed coordinates $X_{i,j}$ are generated. The potential log-IV $Y_{i,j}$ is set by:

$$Y_{i,j} = m_0(X_{i,j}) + \sum_{l=1}^{3} \beta_{i,l} m_l(X_{i,j}).$$

We choose $L = 3$ similarly to the data analysis from Section 5.2. The invariant basis function m_0 we keep 0. The dynamic functions m_1-m_3 are orthogonal planes:

$$\begin{aligned} m_1(u_1, u_2) &= 1, \\ m_2(u_1, u_2) &= -5u_1 + 5, \\ m_3(u_1, u_2) &= -2u_2 + 1, \end{aligned}$$

where u_1 is the moneyness coordinate and u_2 is time to maturity coordinate. The loading time series β_i are generated as autoregressive process, but $\beta_{i,1}$ has significantly bigger magnitude than two remaining time series. This choice of the factor functions

$$\begin{array}{c} \\ \beta_1 \\ \beta_2 \\ \beta_3 \\ \widehat{\beta}_1 \\ \widehat{\beta}_2 \\ \widehat{\beta}_3 \end{array} \begin{pmatrix} 1.00 & x & x & 1.00 & x & x \\ & 1.00 & x & x & 0.29 & 0.94 \\ & & 1.00 & x & 0.96 & -0.26 \\ & & & 1.00 & x & x \\ & & & & 1.00 & x \\ & & & & & 1.00 \end{pmatrix}$$

Table 5.4: The correlation matrices of factor loadings β_i and estimated factor loadings $\widehat{\beta}_i$. Only important quantities are presented.

and loading time series is motivated by the results of previous section. We try to mimic the behavior of the IVS by choosing the first function flat and the first time series much bigger than others.

In Figure 5.8 we present the estimated time series $\widehat{\beta}_l$ together with the original time series β_l. The estimated first time series is similar to generated time series β_1. It can be also confirmed by the correlation between these time series as it is shown in Table 5.4, where the correlation between β_l and $\widehat{\beta}_l$ is 1. It is however not the case for the other factors. The second and third time series (β_2, β_3) are different from the second and estimated loadings time series respectively ($\widehat{\beta}_2$, $\widehat{\beta}_3$). Although the resemblance can be found if one swaps the order of the factors, which could be seen in Table 5.4 but the $\widehat{\beta}_2$ and β_3 are not perfectly correlated.

This issue can be also illustrated in Figures 5.9-5.10 where the original and estimated functions are plotted. The invariant and first factor functions shows similarity with the equivalent estimated functions. Pairs (m_2,$\widehat{m}_3$) and (m_3,$\widehat{m}_2$) show some resemblance but the magnitudes of the functions are different. However the fit of the estimated model is almost perfect since the explained variance defined in (3.20) is very close to 1 and Ξ_{AIC_1} and Ξ_{AIC_2} are very close to 0.

Although this simulation study is done only for one case and cannot be generalized it shows that DSFM has potential to estimated correctly the factor function and loading time series. However this study needs to be largely extended which is beyond the contents of this work. Our aim is only show that it is feasible.

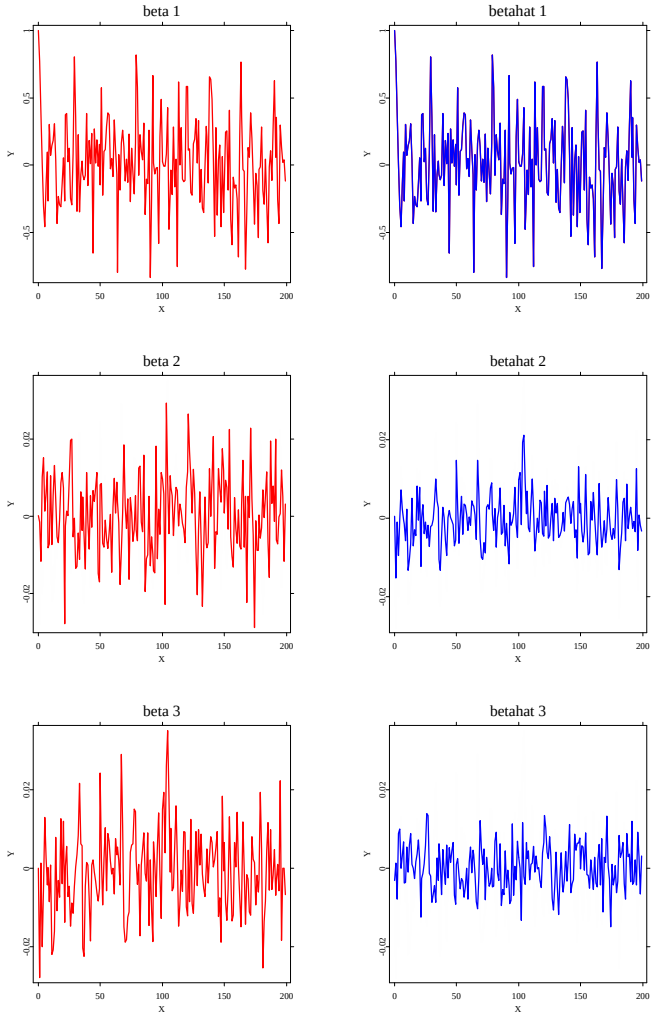


Figure 5.8: Time series of loadings. Left panels: simulated time series β_1, β_2 and β_3. Right panels: estimated time series $\widehat{\beta}_1$, $\widehat{\beta}_2$ and $\widehat{\beta}_3$.

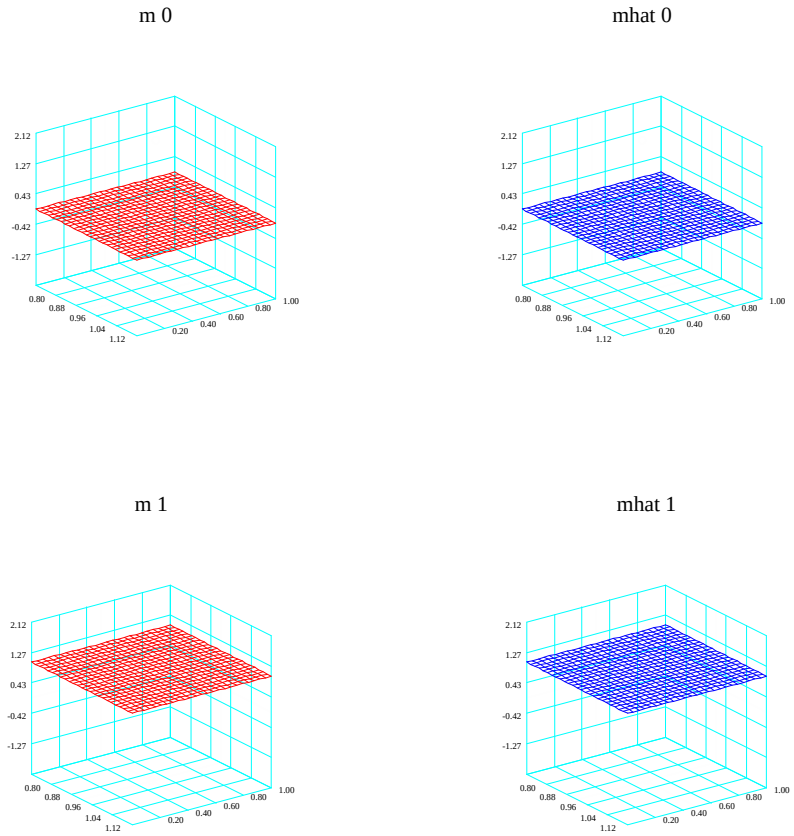


Figure 5.9: Invariant basis functions and first dynamic basis functions. Left panels: true functions m_0 and m_1. Right panels estimated functions $\widehat{m}_0$, $\widehat{m}_1$.

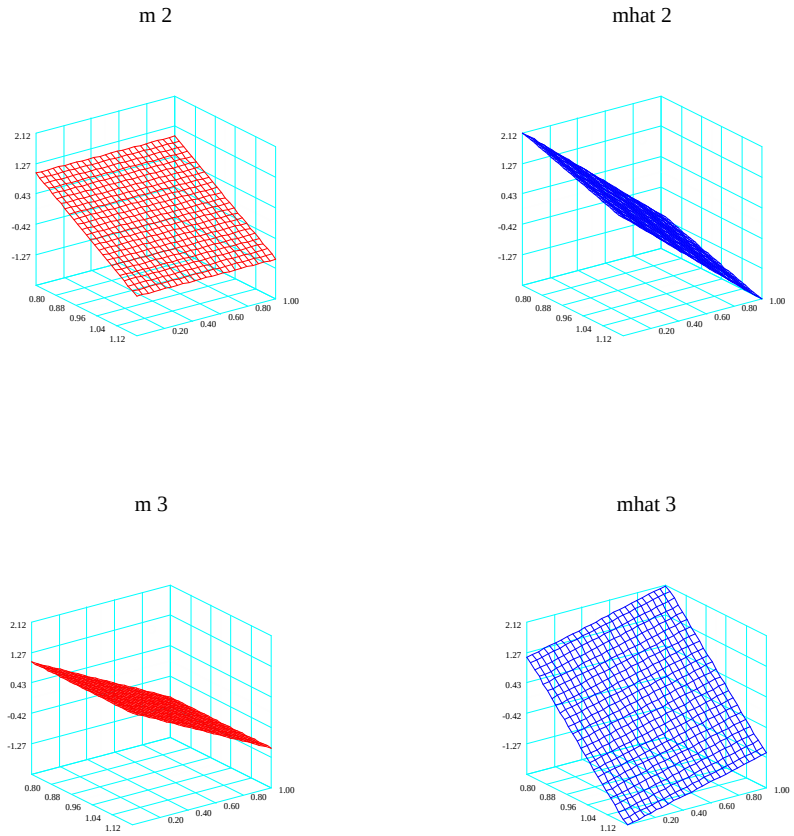


Figure 5.10: Second and third basis functions. Left panels: true functions m_2 and m_3. Right panels estimated functions $\widehat{m}_2$, $\widehat{m}_3$.

5.4 Hedging exotic options

The regulated derivative markets allow to trade not only the plain vanilla options but recently also the barrier options become wide traded financial instrument. The barrier option is a path dependent option, which mean that payoff depends not only on the asset price in expiry time but also on the asset's price in the life time of the option. The payoff of the option is defined in the similar way like the payoff of plain vanilla options but it will be the fact only if in any time before the expiry the price hits or cross the pre-specified level (knock-in) or just opposite if the barrier is not reached (knock-out). One may classify the barrier options according to payoff type (put/call), position (long/call), condition on the payoff (knock-in/knock-out), the relation of the underlying and barrier (up - if the price is below the barrier / down - if the price is above the barrier).

In the BS model from Section 2.1 one may derive analytical prices for the barrier options. Similarly to (2.4) the price is dependent on unknown parameter σ.

The natural question, which arise is how to hedge the short position in barrier option. It is particularly important for down-and-out put and up-and-out call. If the price is near the barrier level the option resembles in some way digital options by yielding *a*ll or nothing. The next feature of these options is their negative relation to volatility σ. It is opposite to the plain vanillas, which have positive relation to σ. If the chance of hitting the barrier increases than the price of the option decreases.

The first approach to hedge the knock-out barrier option is construct the portfolio which replicate its payoff if the barrier is not reached and have no value when the underling is equal the barrier. One have to assume unrealistic trading conditions (like possibility of buying any fraction of the asset) not only for underlying but for all possible vanillas too. As an example consider the knock-out call option (C^{KO}) with strike 100 and barrier 90. Assume additionally that the underlying has no drift and the IV on the market depends on time not on strike. The hedge portfolio can be constructed from one position in plain vanilla call with strike 100 and 100/90 short put with strike 81. If the barrier is not hit up to maturity the payoff of the knock-out call is same like the payoff of the plain vanilla call and the put is worthless. If the underlying is 90 then the value of the put and call in the hedge portfolio are the same. It is independent on the σ, however the interest rate r has to be 0.

The presented static hedging requires many strong assumptions and may not be applied in practice. More feasible approach is a dynamic hedging. For each particular time t the barrier option is replicated with a certain amounts of underlying and call option (C). These amounts can be calculated with greeks, which may be easily computed, since the price is given analytically. In practice however the change of the barrier option value may not be fully hedged with other financial instruments. It can be only

approximate up to some level and the approximation can be given as:

$$\Delta C^{KO}(\Delta S, \Delta\sigma) \approx \frac{\partial C^{KO}}{\partial S}\Delta S + \frac{\partial C^{KO}}{\partial \sigma}\Delta\sigma.$$

Here we present only the change of the barrier option value with respect to asset value S and volatility σ. The hedge portfolio HP consist of a_1 units of underlying and a_2 call options. The sensitivity of the hedge portfolio $HP = a_1 S + a_2 C$ with respect to S and σ should be equal to the sensitivity of the C^{KO}. The hedge coefficients $a1, a2$ are given by the equation:

$$\begin{pmatrix} 1 & \frac{\partial C}{\partial S} \\ 0 & \frac{\partial C}{\partial \sigma} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial C^{KO}}{\partial S} \\ \frac{\partial C^{KO}}{\partial \sigma} \end{pmatrix}.$$

First the joint portfolio (exotic + hedge portfolio) has to be vega neutral and afterwards the delta hedge is adjusted.

The presented hedging strategy however disregard the possible changes of the IVS. We know that the IVS changes the shape throughout the time. The other problem of BS approach is the single parameter σ, which explains the volatility. Although to price non observable plain vanilla options consistently with market prices it is enough to interpolate IV from traded options with the same moneyness it is not the case for the barrier options. Their price depends not only on the classical moneyness but also on the relation between barrier and the spot price.

In order to price the barrier options consistently with the smile LV model can be employed. One have to solve the PDE (2.13) with modified boundary conditions for barrier case. In this approach the greeks delta $\frac{\partial C^{KO}}{\partial S}$, gamma, $\frac{\partial^2 C^{KO}}{\partial S^2}$ and theta, $\frac{\partial C^{KO}}{\partial t}$ can be read from finite difference scheme. But what about vega $\frac{\partial C^{KO}}{\partial \sigma}$? It is no longer the single number but the whole surface.

The first natural proposal is classical vega shifts. One simply calculates the sensitivity of up-and-down movements of the whole IVS. It explain significant part of the IVS variation but still does not yield any protection to the IVS shape changes. Extension to the classical vega is a bucket vega hedging where one calculates sensitivity with respect to the changes of each particular maturity. Still the smile risk remains not hedged. One may also try super-bucket vega hedge by calculating the sensitivity with respect to parallel changes of the small parts of IVS in one maturity or even for each grid point. Then all possible smile changes are hedged but it yields infeasible many parameters and is practically impossible.

To overcome the problem with to many hedging parameters one has to assume a model for IVS dynamics, which additionally reduce the dimension. The option prices

are given from BS formula (2.4) and then one obtains the LVS by applying (2.12). The natural choice the model is DSFM. The whole dynamics is describe with L factors. Thus one has to calculate only sensitivity to the dynamic factor functions. The first function $\widehat{m}_1$ reflects up-and-down dynamics, the second $\widehat{m}_2$ term structure and third one $\widehat{m}_3$ smile dynamics. We recall the vega hedging example from Fengler et al. (2005) for one particular barrier option. It means that we do not consider the term structure effects, which has to be taken into account for the portfolio containing more options with different maturities.

In order achieve vega neutral joint portfolio $(-C^{KO} + HP_{vega})$ consider the hedge portfolio $HP_{vega} = a_1 HP_1 + a_2 HP_2$. The HP_1 should be sensitive to β_1 and relatively independent to β_3 changes. It may consist of at-the-money call option. The HP_2 need to have opposite sensitivity, small to β_1 and relatively large to β_3. It may constructed as risk reversal. Risk reversals are combinations of short put with strike K_1 and long call with strike $K2 > K1$ or long put with strike K_1 and short call with strike K_2. They are relatively independent to parallel shifts of the IVS but reflects significant dependence to smile changes. The coefficients a_1 and a_2 are given by the solution of:

$$\begin{pmatrix} \frac{\partial HP_1}{\partial \beta_1} & \frac{\partial HP_2}{\partial \beta_1} \\ \frac{\partial HP_2}{\partial \beta_3} & \frac{\partial HP_2}{\partial \beta_3} \end{pmatrix} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} \frac{\partial C^{KO}}{\partial \beta_3} \\ \frac{\partial C^{KO}}{\partial \beta_3} \end{pmatrix}.$$

Bibliography

Avellaneda, M., and Zhu, Y. (1997). An E-ARCH Model for the Term-Structure of Implied Volatility of FX Options, *Applied Mathematical Finance*, **4**: 81–100.

Bates, D., (1996). Jump and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options, *Review of Financial Studies*, **9**: 69–107.

Black, F., and Scholes, M. (1973). The pricing of options and corporate liabilities, *Journal of Political Economy*, **81**: 637–654.

Borak, S., Detlefsen, K., and Härdle, W. (2004). FFT based option pricing, in Čížek, P., Härdle, W., Weron, R. (eds.) *Statistical Tools for Finance and Insurance*, Springer, Berlin.

Borak, S., Fengler, M. and Härdle, (2005). DSFM fitting of Implied Volatility Surfaces, SFB 649 Discussion Paper, Hümboldt-Universität zu Berlin.

Cont, R., and da Fonseca, J. (2002). Dynamics of implied volatility surfaces, *Quantitative Finance*, **2**: 45–60.

Detlefsen, K. (2005). Hedging Exotic Options in Stochastic Volatility and Jump Diffiusion Models *Master thesis*, Hümboldt-Universität zu Berlin.

Dupire, B. (1994). Pricing with a smile, *Risk*, **7**: 18–20.

Fengler, M. (2004). Semiparametric Modelling of Implied Volatility *PhD thesis*, Hümboldt-Universität zu Berlin.

Fengler, M., Härdle, W. and Schmidt, P. (2002). The Analysis of Implied Volatilities, in Härdle, W., Kleinow, T., Stahl, G., (eds.) *Applied Quantitative Finance*, Springer, Berlin..

Fengler, M., Härdle, W. and Villa, P. (2003). The Dynamics of Implied Volatilities: A common principle components approach, *Review of Derivative Research*, **6**: 179-202.

Fengler, M., Härdle, W. and Mammen, E. (2003). Implied Volatitlity String Dynamics, SFB 649 Discussion Paper, Hümboldt-Universität zu Berlin.

Gijbels, I., and Mammen, E. (2002). Local Adaptivity of Kernel Estimates with Plug-in Local Bandwidth Selectors, *Scandinavian Journal of Statistics*, **25**: 503–520.

Hafner, R. (2004). Stochastic Implied Volatility *Springer-Verlag*, Berlin Heidelberg.

Hafner, R. and Wallmeier, M. (2001). The Dynamics of DAX Implied Volatilities, *International Quarterly Journal of Finance*, **1**: 1-27.

Heston, S., (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Review of Financial Studies*, **6**: 327-343.

Hull, J. and White, A. (1987). The pricing on option on assets with stochastic volatilities, *Journal of Finance*, **42**: 281–300.

Merton, R., (1976). Option pricing when underlying stock returns are discontinuous, *J. Financial Economics*, **3**: 125-144.

Press, W., Teukolsky, S., Vetterling, W. and Flannery, B. (1992). Numerical Recipes in C, *Cambridge University Press*.

Skiadopoulos, G., Hodges, S. and Clewlow, L. (1999). The Dynamics of S&P 500 Implied Volatility Surface, *Review of Derivatives Research*, **3**: 263–282.

Stein, E. and Stein, J. (1991). Stock price distribution with stochastic volatility: An analytic approach, *Review of Financial Studies*, **4**: 727–752.